A review of video-based human activity recognition: theory, methods and applications

Article *in* Multimedia Tools and Applications · July 2024

DOI: 10.1007/s11042-024-19711-w

6 authors, including:

Tanvir Fatima Naik Bukht
Air University

Hameedur Rahman
Air University

Momina Shaheen
University of Management and Technology

Nouf Abdullah Almujally
Princess Nouraa bint abdulrahman University

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-19711-w

A review of video-based human activity recognition: theory, methods and applications

Tanvir Fatima Naik Bukht[1] · Hameedur Rahman[1] · Momina Shaheen[2] · Asaad Algarni[3] · Nouf Abdullah Almujally[4] · Ahmad Jalal[1]

Received: 16 January 2024 / Revised: 25 May 2024 / Accepted: 18 June 2024

Abstract
Video-based human activity recognition (HAR) is an important task in many fields, such as healthcare monitoring, video surveillance, and sports analysis. This review paper aims to give an in-depth look at the current state of the art in HAR from 2018 to 2024. This will include a discussion of the different methods and models used for extracting, representing, and classifying human actions from video, as well as the challenges and limits of this field. The paper will also discuss recent improvements and plans for making HAR systems more accurate and useful. Even though there has been a lot of progress, a few knowledge gaps still need to be filled to make recognition more accurate and efficient. The purpose of this review paper is to offer scholars and professionals an overview of the theory, methods, and applications of HAR in videos. Through a critical analysis of the extant literature, this paper seeks to identify prospective avenues for future research and contribute towards advancing HAR systems that are more precise and efficient. By showing the different ways that HAR can be used, the paper shows how important this field is in many different areas.

Keywords Human activity recognition · Modes of activities · Learning methods · ANOVA

1 Introduction

Technology advancements and data from sensors and CCTV have made it possi- ble to identify routine human behavior and detect anomalies for surveillance pur- poses. These advancements have greatly improved the efficiency and effectiveness of surveillance systems, allowing for real-time monitoring and quick responses to potential threats. Additionally, the integration of artificial intelligence algorithms has further enhanced the accuracy of anomaly detection and human action recognition, making surveillance systems more reliable than ever before [9, 23, 44, 179]. HAR,or Histogram of Oriented Gradients, is a computer vision problem used for classifying objects based on their visual appearance, aiding in object tracking, face detection, and activity recognition. It could be used in healthcare [91], surveillance, sports [134], elderly care [194, 207], and human-computer interaction (HCI) [19, 42, 51]. Lighting, background, crowded

Extended author information available on the last page of the article scenes, the camera's perspective, and the activity's complexity all impact HAR's accuracy. HAR applications have also played a crucial role in emergency response systems, enabling faster and more accurate assistance during critical situations [20].

An intelligent video surveillance system can find unusual things and things that don't belong, like weapons in places where they aren't allowed or things that were left behind. The movie has ambiguous, exceptional, undescribed, scarce, infrequent, surprising, normal, and non-dictionary irregularities [32]. Crowd modeling, tracking, dense projection, counting, and interpreting crowd behavior are all aspects of automated crowd analysis that can help humans discover risks and anomalies

The general process includes monitoring actions, determining features, and spotting out-of-the-ordinary actions [29]. Sensors, machine learning algorithms, and wireless connections have made developing novel medical and assistive technology systems feasible. The program has enhanced social inclusion and participation among older adults, promoting a sense of belonging and community. Sensor-based HAR uses the five stages of sensor selection, data collection, feature extraction, model training, and model testing to learn actions from a sequence of observations [115]. It is critical to detect abnormal conduct because it might vary based on the circumstances, making it difficult to describe. Various methodologies are employed to detect abnormal behavior [32]. Surprises, deviations, criminals, or irregularities play as important a role in literature as do anything else. Human-intelligent video systems can easily recognize the weapons and lost items, whereas automatic crowd analysis better public safety and security by identifying possible risks and responding on the time. Sophisticated algorithms and AI technologies reinforce prevention and intervention methods, safeguarding the integrity of people who are in crowded spaces. [170]. Healthcare systems have revolutionized the industry by enabling remote patient monitoring, personalized treatment plans, and improving the quality of life for disabled individuals. Assistive technologies enhance mobility, streamline administrative processes, and reduce paperwork. Continuous learning and analysis of patterns provide real-time alerts and interventions, ensuring individual well-being [36, 164, 195, 206].

The major contributions of our paper are as follows.

- We explain a comprehensive study that involved different modes of activities, applications, user activities, and strategy-based activities.
- Evaluated various learning techniques and models that we used in our research, such as data-driven and knowledge-driven.
- We discuss detailed data sources and show how we acquired the data for the sake of reliability as well as clarification.
- The paper covers a brief overview of the latest developments in HAR as well as the various applications, challenges, and possible future improvements.

The paper is structured as follows: Materials and methods are represented in Section 2 , along with research questions and PRISMA, and human activity taxonomy is defined in Section 3. Section 4 represents modes of activities and their corresponding features, while Section 5 represents learning methods and algorithms used in the study. Section 6 represents data sources and data collection procedures. Section 8 focuses on the study's challenges and limitations. and the section 9 is about the conclusion.

2 Motivation & objectives

The databases at Web of Science were utilized to compile a collection of HAR research articles. The collected articles were then filtered based on relevance to the field of interest, resulting in a final dataset of 177 research articles. The selected articles cover many topics, including HAR algorithms, sensor technologies, machine learning techniques, and applications in various domains such as sports, healthcare, and smart homes. We also looked into other databases, but their access was restricted due to subscription requirements or a lack of availability. The articles were sourced from academic databases and online archives using relevant keywords and phrases. Multiple sources were consulted to ensure a diverse collection. Boolean operators and filters were applied to refine the results, narrowing them down to recent publications. We use the keywords ”human activity recognition” AND (”types” OR ”mode ”) AND (”applications” OR ”user activities” OR ”Suspicious” OR ”classification”) and another string ”human activity recognition” AND (”activity classification” OR ”mode classification”) for the first question. To answer the second question, we are using the following keyword strings: ”human activity recognition” AND (”theories” OR ”theory” OR ”learning methods” OR ”machine learning” OR ”deep learning” OR ”algorithms”) AND (”Supervised learning” OR ”unsupervised learning”). We are using the following keyword combinations to address the third question ”human activity recognition” AND (”data sources” OR ”sensor data” OR ”datasets”) AND (”challenges” OR ”issues” OR ”limitations” OR ”data collection” OR ”data analysis”) were used as the keywords to search for relevant research papers in the field of computer vision. These key- words cover various topics related to understanding human actions, interactions, and behaviors from visual data. Using these keywords, researchers can explore various aspects of activity recognition, including detecting specific actions or activities, analyzing group behavior, identifying abnormal or suspicious behavior, and detecting violent events. These research areas have applications in surveillance systems, video understanding, HCI, and many other domains.

2.1 Research questions

Table 1 provides an overview of the research questions that guide the study, highlighting the underlying motivations behind these questions. It also presents potential solutions or approaches that can be explored to address these research inquiries. By organizing this information in a structured manner, Table 1 offers a concise and comprehensive understanding of the research objectives and the corresponding avenues for investigation Tables 2, and 3.

Initially, 2243 articles were collected using a keyword search on a popular academic database. The articles were then screened based on their relevance to the research topic, resulting in a final selection of 177 articles for further analysis. The selection process involved carefully screening the abstracts and titles of each article to ensure they aligned with the research objectives. Figure 1 shows the steps performed for the selection of articles, which are identifying the research question or topic, conducting a thorough search using databases and search engines, applying inclusion and exclusion criteria, assessing article quality and relevance, extracting relevant data, analyzing findings, interpreting results, and summarizing the finding.

Table 1 Discussing research questions, the motivation behind them, and the answers

Research Questions	Motivation	Answers
1. How are different modes of activity classified, and what are the potential applications of HAR technology in various fields?	To understand and anal- yse the categorization of different modes of activity and explore the wide- ranging applications that HAR technology holds	Section 4, Tables 4, 5, 6, 7, 8 and 9
2. What are the current theo- ries and learning methods used in HAR?	The motivation behind this is to understand the vari- ous approaches and Methods implemented in HAR comprehensively	Section 5, Table 12, Figure 8
3. What are the various data sources used in HAR research, and what challenges do re- searchers face in collecting and analyzing this data?	Explore the different types of data sources used in HAR research and chal- lenges for better under- standing	Section 6 and 8, Table 15

Table 2 Summary of single-factor summary for modes of activities

Groups	Count	Sum	Average	Variance
Types of Activities	2	40	20	50
Applications	3	20	6.667	5.33
User Activities	2	17	8.5	0.5
Strategy based Activities	2	17	8.5	12.5

Table 3 One-way ANOVA

Source-of-Variation	SS	df	MS	F	P-value	F-crit
Between Groups	240.6	3	80.1	5.5	0.0494	5.409
Within Groups	73.67	5	14.74			
Total	314	8				

3 HAR taxonomy

Human activity is a series of actions by one or more individuals. HAR uses advanced techniques like machine learning and sensor data analysis to develop intelligent systems for improved monitoring and assistance [47, 80, 103, 202, 205, 208]. The range of activities encompasses indoor and outdoor settings, with indoor options including sedentary activities like sitting and lying down and ambulatory activities like walking. On the other hand, outdoor activities encompass more physically engaging pursuits such as playing football or horseback riding. HAR is currently employed across various application domains in the existing literature [61, 79]. This study categorises the modes of activity in HAR into four distinct categories: types of various actions, applications, user, and strategy-based activities. There are three primary categories of activities: static, which involves maintaining a fixed position, such as standing or sitting; dynamic, which involves movement, such as walking or running; and interactive. The application-related research activities can be divided into healthcare, suspicious activities, and surveillance. The studies referring to user activities are divided into two distinct categories: single and group activities. It is important to note that most of the studies in this research fall into the categories of user actions and daily living applications. Methods for learning HAR can be broadly divided into two categories: data-driven and knowledge-driven.

The nature and accessibility of data are essential considerations in the field of HAR. The researchers utilize various HAR data from various video and sensor- based sources. The present study employed a systematic approach to categorize the extant body of literature according to the nature of the data utilized, including video and sensors. The data obtained through vision-based methods is subsequently categorized according to its nature, which may include video-based and sensor-based methods. The literature on HAR incorporates a variety of video sources, including footage from closed-circuit television (CCTV), smartphones, Kinect devices, and YouTube. Conversely, video-based HAR for mobile applications relies on data from social media platforms and camera images.

Existing HAR studies' open challenges and limitations are classified into seven categories: data collection, segmentation, data preprocessing, feature extraction, hardware and techniques, complex activity detection, and activity misalignment. Video-based data is more substantial and demanding of processing power than sensor-based data. Figure 2 illustrates the overall taxonomy of existing HAR literature.

4 Modes of activities

This study categorized the existing HAR literature into four categories: type of activities, applications, user activities, and media activities, as mentioned above. The following sections provide in-depth analyses of every category.

4.1 Types of activities

Types of activities are divided into subcategories: static, dynamic, and interaction activities. Some static and dynamic datasets sucg as UCF ARG [112] , ASLAN [81], Sport-1M [70], Charades [152] , DALY [178], MultiTHUMOS [193] , AViD [128] , AVA [46] , Charades-Ego [153] , HA500 [28] . HVU [34], Kinetic 700 2020

[156] Static activities do not involve much physical movement and are usually done in a stationary position, such as yoga, sitting, standing, and weightlifting. On the other hand, dynamic activities involve physical movement and usually require more energy, such as running, swimming, or dancing. Interaction activities focus on social engagement and communication and can be human-to-human or human- to-object activities. These subcategories help to provide a better understanding of the different types of activities.

4.1.1 Static and Dynamic Activities

Static activities refer to situations in which an individual remains static in re- lation to the configuration of the surrounding environment. The static activities observed in our environment include standing still, sitting on a chair, sitting on the ground, and lying on the ground. Dynamic activities refer to actions in which an individual exhibits continuous movement in relation to the configuration within the surrounding environment.

Researchers have identified actions based on pre-segmented data to implement a robust HAR system, despite the challenges it presents. The HAR from Najeh et al. [113] finds activities that happen over time and over again by comparing their correlation and F1 scores, which were between 0.63 and 0.99 in a real-world CASAS case study.

Khodabandelou et al. [77] introduces a fuzzy logic-based deep learning algorithm that predicts lower limb exoskeleton users' daily activities using real-time locomotion data, estimating gait mode transitions and evaluating performance using dynamic data.

K¨oping et al. [84] developed a smartphone-based SVM-based framework for real-time activity identification that achieved 87.1% accuracy using extracted features. Reliability was enhanced using KPCA and LDA. In their study, Manzi et al. [102] present an activity recognition system using depth camera skeleton data and machine learning techniques. It classifies actions based on postures using multiclass SVM and X-means algorithms. The method outperforms state-of-the-art methods with only 4 seconds of input data. In their study, Kellokumpu et al.

[71] A novel method detects human activity using dynamic texture descriptors, simplifying computation. It works with picture data and compares results to the best methods, utilizing computer vision research advances. Shelke and Aksanli's

[150] method efficiently implements smart spaces using low-resolution data and is trained using Naive Bayes (NB),Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT) , and Artificial Neural Networks (ANN), achieving a 99.96% accuracy rate in continuous HAR.

Ahmad et al. [4] introduced a hybrid feature selection approach that optimizes performance on resource-constrained hardware using filter and wrapper methods, achieving 96.7% accuracy rate while overcoming challenges like annotated data, computational expenses, and system resource demands. In their study, Khan and Ahmad [76] introduced a novel attention-based multihead model that excels on UCI HAR and WISDM datasets with three convolutional heads. Table 4 provides a summary of the research that has been performed on static and dynamic activities.

The data from a one-way analysis of variance (ANOVA) comparison of four groups of types of activities, applications, user activities, and strategy-based activities is shown in the table. The count, sum, average, and variation for each group are detailed in the "Summary" section. The type of activities group's statistics are as follows: 40 totals,2 counts, 20 averages, and 50 variances Similar results can be seen for the applications group, which has a count of 3, a sum of 20, a variance of 5.334 and an average of 6.667, and so on, as shown in the tables below. These statis- tics provide valuable insights into the performance and behavior of each group. By analyzing the counts, totals, averages, and variances, we can better under- stand the activities and applications within each group. This information can be used to identify trends, patterns, and areas for improvement, ultimately leading to more informed decision-making and enhanced productivity. The ANOVA part provides a comprehensive analysis of the sources of variation, including degrees of freedom (df), sum of squares (SS), F-ratio (F), mean square (MS), p-value, and critical F-value (F crit). This information allows for a deeper understanding of the statistical significance and relationship between the variables in each group. The study's p-value, below the accepted significance threshold of 0.05, of 0.0494 indicates a significant difference in averages between the four groups. The 5.442 F-ratio demonstrates that there is a significant difference between the groups. The crucial F-value is 5.409 as shown in Tables 2, 3

The existing literature on HAR primarily focuses on analyzing the frequency of different modes of activity. Specifically, it reveals that 37% of the activities studied are dynamic and static, while 63% involve interaction. Additionally, when it comes to applications, 40% are considered suspicious, 40% are related to surveillance, and 20% pertain to healthcare. Furthermore, user activities are categorized into single user activity 47% and grouped user activity 53%. Lastly, strategy-based activities are divided into offline 65% and 35% streaming as shown in Figure 3.

4.1.2 Interaction activities

This session aims to comprehensively analyse interactions in different contexts and contribute to understanding complex systems. The single human activity and human-to-human interaction dataset is a comprehensive collection of various individual activities and interactions. Human to human interaction datasets may include BIT-interaction [83], UT-interaction [138] , HMDB51 [85], and UCF101

[157] etc.

Table 4 A comprehensive overview of the existing research on dynamic and static activities

Ref	Year	Description
Kellokumpu et al. [71]	2008	New method detects human activity using dynamic texture descriptors, simplifying computation and comparing results
Manzi et al. [102]	2017	Present an activity recognition system using depth camera skeleton data and machine learning techniques. It classifies actions based on postures using multiclass SVM and X-means algorithms. The method outperforms state-of-the-art methods with only 4 s of input data
K¨oping et al. [84]	2018	The proposed data integration framework combines data collection and a codebook-based feature learning method, utilizing a non-linear SVM for minimalization
Alghyaline et al. [6]	2019	The proposed real-time human activity detection using the Kalman Filter, ahomography, and YOLO object detection
Shelke et al. [150]	2019	The study used various machine learning algorithms to train models on low-resolution using Logistic Regression, Naive Bayes, SVM, Decision Trees, Random Forests, and ANN
Ahmad et al. [4]	2020	The sequential Forward Floating Selection SFFS extracts features, and SVM classifies activities in a hybrid feature selection process
Khan et al. [76]	2021	A one-dimensional CNN framework with three convolutional heads was proposed. This framework improves representation and automates feature selection
Singh et al. [155]	2021	This work proposes a ConvNet model using RGB frames and Bi-LSTM for HAR, achieving the best classification accuracy on common datasets
Ishikawa et al. [64]	2021	The ASRF framework uses an ASB to categorize video frames, a BRB to estimate action boundaries, and a loss function to smooth action probabilities
Najeh et al. [113]	2022	Researchers used temporal correlation identification to deter- mine if the current action is a continuation of a previous activity or novel
Khodabandelou et al. [77]	2023	Fuzzy logic deep learning algorithm improves lower limb exoskeleton user activity sequences and gait estimation
Helmi et al. [55]	2023	This study develops a robust HAR system combining deep learning and swarm intelligence. It compares three binary variants and finds the one with the best performance

Table 5 comprehensive overview of the existing research on interaction activities

Ref	Year	Description
Shehzed et al. [149]	2019	Multi-person tracking system detects normal/abnormal events with 88.7% accuracy and 95.5% detection rate
Kim et al. [78]	2019	Study propose a video-based HAR system for elderly monitoring, recognizing daily activities in indoor environments using skeleton joint features
Nadeem et al. [111]	2020	Study uses linear discriminant analysis and artificial neural network for precise human action detection in KTH and Weizmann datasets
Jalal et al. [66]	2020	A study presents a video-based HAR system for elderly monitoring, achieving 91.25% accuracy on various datasets
Pervaizet al. [127]	2021	New visual surveillance approach uses Gaussian filter, background removal, skin verification, body point detection, centroid of silhouettes, and jacquard similarity index
Alarfaj et al. [5]	2022	Proposed system for human object interaction recognition achieves 87.5% accuracy on the MPII dataset
Hartmann et al. [53]	2022	Present interactive real-time HAR uses hidden markov models, enabling users to engage, test performance, and extend detected classes
Ghadi et al. [43]	2022	Parts-based model recognizes complex human-object interactions in aerial images using gamma correction, denoising, and Felzen-szwalb's algorithm
Tang et al. [165]	2022	Study introduces Dual-branch Interactive Network (DIN) for handling multi-channel time series in HAR, combining CNN and Transformer advantages
Mahwish and Jalal [126]	2023	Project improves visual classification and event analysis using pre-processing, feature extraction, optimization, and artificial neural networks
Usman et al. [11]	2023	Research article introduces a drone-based system for human recognition, outperforming existing methods with 80.03%, 48.60%, and 78.01% accuracy
Tanvir fatima et al.[24]	2023	This article presents action recognition techniques using decision trees, utilizing HSI color transformation, filters, feature extraction, shape and texture extraction, vectors, and t-SNE for classification

Table 6 Human activity recognition datasets

Ref	Category and Dataset name	Classes no	year	Resolution
Human—human Interaction				
Ivan et al. [86]	HOHA	12	2008	
kong et al. [83]	BIT-interaction	8		
Ryoo et al. [138]	UT-interaction	6	2010	720×480
Kuehne etal [85]	HMDB51	51	2011	320×240
Soomro et al. [157]	UCF101	101	2012	320×240
Barekatain et al. [16]	Okutama- Action	10	2017	3840×2160
Human Object Interaction				
Soomro et al. [137]	UCF Sports		2008	720×480
Niebles et al. [116]	Olympic Sports	16	2010	
Oh et al. [120]	VIRAT	23	2011	1,920×1,080
Reddy and Shah [135]	UCF50	50	2013	
Barman et al. [17]	Games action dataset		2018	1,080p
sultani et al. [162]	Youtube Aerial dataset	8	2021	

Human-to-human interaction includes dynamic exchanges and social interactions. The BIT and UT Interaction datasets are commonly used in human-human interaction recognition research [184]. The BIT Interaction dataset offers a comprehensive understanding of human communication through face-to-face conversations, gestures, and body language, while the UT Interaction dataset focuses on group interactions. These datasets contribute to robotics, virtual reality, and artificial intelligence advancements by enhancing the understanding of human behavior and improving interactive system design. Recognition plays a vital role in enhancing these technologies. The BIT-Interaction dataset features eight human interactions, including bowing, boxing, handshake, high-five, embrace, kick, and pat, captured in a realistic environment with partial occlusion, varying sizes, and illumination variations. The UT-Interaction dataset comprises six human interaction classes: push, kick, hug, point, and punch. It includes two videos, UT-set-1 and UT-set-2, captured in distinct environments Figs. 4, 5, and 6. This diverse range of human interactions makes it a valuable resource for training and testing interaction recognition models [138].

The interaction between humans and non-human entities is referred to as human-to-object interaction. We also analyze the VIRAT 1.0 Ground dataset and the VIRAT 2.0 Ground dataset in our investigation of interaction [30]. These datasets are focused on human-vehicle interaction tasks, demonstrating the wide range of interactions that can be investigated. We emphasize the adaptability and practical consequences of evaluating interactions within complex systems by including object interaction in our research. Understanding object interaction is critical because it allows us to learn how humans and non-human entities coexist and collaborate. The VIRAT 1.0 Ground dataset contains rich information about human-vehicle interactions, allowing researchers to investigate pedestrian crossing, vehicle avoidance, and traffic congestion scenarios. Similarly, the VIRAT 2.0 Ground dataset extends this by incorporating more complex interactions like item manipulation and tool usage.

The VIRAT 1.0 and VIRAT 2.0 Ground datasets analyze six human-object vehicle interactions, including trunk closing, opening, loading, unloading, entering, and exiting. These datasets consist of 3 h and 8 h of video in parking lot backgrounds, with two categories of activities: single-object and two-object. The UCF50 was established in 2012 by the computer vision research institute of the University of Central Florida [135]. This project's theme is that it consists of 50 action classes that were all taken from real YouTube videos [85]. provides an overview of HMDB, which is compiled from various sources, primarily movies and a small amount from open databases like YouTube, Prelinger Archives,and Google Videos. The dataset comprises 6849 clips split into 51 action categories, each with at least 101 clips. The HMDB dataset is widely used in computer vision research for action recognition and video analysis, providing a diverse collection of video clips for algorithm development and evaluation in real-world scenarios.

4.2 Applications

Application of HAR techniques such as surveillance [13, 99, 122], suspicious ac- tivity [68, 72, 143], and healthcare includes detecting abnormal behavior in public spaces, identifying potential threats, and ensuring the safety of individuals. HAR aids healthcare professionals in monitoring patient movements, providing real-time feedback on physical therapy exercises, and adjusting treatment plans.

4.2.1 Suspicious

Suspicious human activity can include behaviors such as frequent visits to restricted areas, unusual purchases of large quantities of chemicals or weapons, and attempts to gain unauthorized access to sensitive information. Reporting any suspicious activity to the appropriate authorities is important to maintain safety and security. In this paper [35] , a real-time system for highly accurate 2D pose estimation and convolutional neural network recognition of suspicious human activity is presented. The system extracts human skeletons from video frames and categorizes them according to actions like trespassing, falling, and fighting. The system generates alerts via alarms, messages, and email to stop unusual activities in hospitals and home security. The proposed method outperformed previous methods on the UCSD, UMN, and Avenue datasets, demonstrating exceptional performance. Shoplifting individuals can easily detach labels from merchandise, even under EAS surveillance. CCTV cameras transmit live video footage to a convolutional neural Network (CNN) model, identifying illicit behaviors like shoplifting, robbery, and unauthorized entry. The CNN model triggers an alarm system, achieving an 89% accuracy rate compared to alternative systems [133]. The proposed method effectively detects shoplifting incidents by utilizing advanced computer vision techniques. It detects subtle behaviors and differentiates between suspicious activities like robberies and break-ins, enhancing store security and preventing potential incidents.

In their study, Jyotsna and Amudha [10] A deep learning methodology was used to accurately classify normal and abnormal activities in video frames in an academic setting. The approach used visual cues like motion, color, and texture to represent activity patterns. A deep neural network was trained on a large dataset to classify activities as normal or abnormal accurately. The accuracy rate was 87.15%. This framework by Khan et al. [148] addresses the need for multiple cameras to effectively detect suspicious human behavior in large and complex areas. The framework can accurately identify unusual and suspicious movements by strategically utilizing video statistics from CCTV cameras placed at constant positions. Additionally, including a widget mounted in indoor environments further enhances the system's capability to promptly trigger alarms when such behaviors are detected.

4.2.2 Surveillance

The system developed by Mahdi and Jelwy [101] utilized video surveillance cameras to monitor academic environments and detect any unusual situations that may require intervention. With an impressive accuracy rate of 95.3%, the system effectively alerted the appropriate authorities in a timely manner. This highlights the significance of video surveillance in enhancing security measures while emphasizing the primary objective of HAR, which is to identify and classify various activities captured in videos.

The proposed CNN model in [39] for multiple action detection, recognition, and summarization on a video dataset achieves 98.9% accuracy in identifying actions. Training on a large and varied dataset allows the model to generalize well to various action scenarios and accurately identify different actions. By training the CNN model on a sizable and varied dataset, it is possible to achieve this high accuracy while also ensuring that it can generalize well to various action scenarios. By comparing the HOG of frames in the TDMap, the model can accurately identify and classify different actions accordingly. This capability makes the proposed CNN model valuable for video analysis and understanding human behavior in various applications, such as surveillance systems or motion analysis in sports.

Qin et al. [130] proposed using video surveillance to detect and prevent criminal activities in retail malls (DPCA-SM). This high level of accuracy demonstrates the effectiveness of the DPCA-SM approach in detecting and preventing criminal activities in shopping malls. By analyzing video footage in real-time, the system can track individuals and identify suspicious behavior, allowing security personnel to respond quickly and effectively. The successful evaluation of both real and private datasets further validates the reliability and potential of this surveillance system for enhancing security measures in public spaces.

The proposed method in [196] achieved high accuracy in accurately identifying and classifying activities in videos captured in various scenarios. By utilizing spatio-temporal cubes as an intermediate concept, the method effectively handled cases with different scales, multiple instances, and large fields of view. The experimental results from different benchmark datasets and challenges showcased the superior performance of the approach, positioning it as a promising solution for activity detection in surveillance and driving applications. The performance of TRECVID ActEV 2020/2021, NIST ActEV SDL UF/KF, CVPR ActivityNet ActEV 2021, and ICCV Road 2021 in various driving and surveillance scenarios was tested. Table 7 provides a summary of the research that has been performed on Surveillance.

4.2.3 Healthcare

By incorporating data from various sensors, the framework can capture various physiological and environmental factors that may impact a patient's health. This holistic approach enables healthcare professionals to make more informed and timely decisions, improving patient outcomes [27, 73, 172]. Gumaei et al. [48] introduced a comprehensive framework that uses multiple sensors and incorporates a hybrid deep learning model. This approach

Table 7 A comprehensive overview of the existing research on Suspicious, Surveillance and Healthcare

Ref	Year	Description
Suspicious		
Rajpurkar et al. [133]	2020	Proposed method outperforms previous methods on UCSD, UMN, Avenue datasets for shoplifting detection using CCTV cameras
Jyotsna and Amudha [10]	2020	Deep learning accurately classifies normal and abnormal activities in academic video frames using visual cues
Khan et al. [148]	2020	The framework uses multiple cameras to detect suspicious human behavior in complex areas. It uses CCTV cameras' video statistics to identify unusual movements and trigger alarms promptly when detected
Dileep [35]	2022	Real-time system uses 2D pose estimation and convolutional neural network recognition to detect human activity, categorize skeletons, and alert hospitals and security
Surveillance		
Mahdi et al. [101]	2021	Proposed method for automatically detecting abnormal behavior in academic contexts using VGG and LSTM networks
Elharrouss et al. [39]	2021	Efficient method for TDMap HOG identifies human actions using CNN model comparing existing and newly-generated HOG
Qin et al. [130]	2021	Proposed DPCA-SM framework for detecting suspicious activity in the retail mall using VGG-trained frames and store scenarios
Lijun et al. [196]	2022	Proposed real-time activity detection system using Argus + + for multi-scale video streams
Healthcare		
Subasi[160]	2018	IoT advances healthcare, enabling automated elderly activity monitoring with 99.89% accuracy using data mining
Uddin and Hassan [167]	2018	Deep Convolutional Neural Network for smart healthcare activity recognition uses body sensor signals, evaluated using Mhealth dataset
Gumaei et al. [48]	2019	They developed a smart healthcare framework with multiple sensors and a hybrid deep learning model for 90% accuracy
Taylor et al. [166]	2020	Real-time motion detection with 96.70% accuracy improve fitness monitoring, geriatric care, and personalized treatment plans

Table 8 User activity recognition datasets

Ref	Category and Dataset name	Actors	Classes	year no	Resolution
Single User					
Singh et al. [154]	MuHAVi	14	17	2010	720×576
Perera et al. [125]	Drone-Action	10	13	2019	
Barekatain et. al. [16]	Okutama-Action	10	10	2017	3840×2160
Weinland et al. [177]	INRIA IXMAS	11	13	2006	
Blank et al. [107]	Weizman	10	9	2005	180×144
Schuldt et al. [146]	KTH	25	6	2004	160×120
Group User					
Shao et al. [147]	WWW Crowd		94	2015	640×360
Kuehne etal [85]	HMDB51		51	2011	320×240
Deng et al. [31]	Hollywood2		12	2009	400×300,300 x 200
Heilbron et al. [26]	ActivityNet	203		2015	1280×720
Abu-El-Haija et al. [2]	YouTube 8 M	4716		2016	
Sultani et al. [163]	UCF Crime	13		2018	

was designed to address the inherent limitations associated with single sensing in the context of smart healthcare environments. The framework processes multimodal data using simple recurrent units (SRUs) and gated recurrent units (GRUs). This results in an accuracy of more than 90% in less than 1.7 seconds. This makes it a promising solution for real-time monitoring and decision-making in healthcare settings.

Uddin and Hassan [167] introduced a Deep CNN architecture that applies signals obtained from various body sensors, including accelerometers, magnetometers, and gyroscopes, to extract relevant features. The application of a deep CNN was implemented in conjunction with Gaussian kernel-based PCA for the purpose of activity recognition in the field of smart healthcare. The efficacy and applicability of the approach are evaluated using the Mhealth dataset in the context of cognitive assistance. IoT is gaining popularity in healthcare, enabling automated daily activity monitoring for the elderly. Subasi[160] presents an intelligent healthcare system using IoT technology and data mining techniques, outperforming competition with 99.89% accuracy.

Taylor et al. [166] showed quasi-real-time human motion detection using a noninvasive approach. They created test scenarios for standing up or sitting down utilizing software-defined radios (SDRs) and the RF algorithm to reach 96.70% accuracy. Medical images are valuable sources of information pertaining to diseases, enabling their real-time utilization for the purposes of disease detection and intervention. This study substantially contributes to the medical domain, particularly in fitness monitoring and geriatric care. In addition, using medical images in real time allows for early detection and intervention, improving patient outcomes. Furthermore, integrating these technologies in fitness tracking and elder care can provide valuable insights into individuals' health status and help tailor personalized treatment plans. Table 7 provides a summary of the research that has been performed on Healthcare.

Table 9 A comprehensive overview of the existing strategy-based activity research

Ref	Year	Approach	Modality	Description
Offline				
Wang et al. [175]	2015	Handcrafted features	Unimodal	The study suggests a way to recognize actions by using optic flow histograms and TDD. This method makes it easier and more accurate to get both spatial and temporal information from action sequences
Mukherjee et al. [110]	2018	Deep learning	Multimodal	ResNet 101 is a deep residual network architecture that generates dynamic images and movies, improving motion capture accuracy and minimizing complexity compared to traditional RGB based methods
Franco et al. [40]	2020	Handcrafted features	Multimodal	Multimodal technique combines skeletal and RGB data to capture human actions, improving accuracy and system performance, enabling robust and accurate systems
Zhang et al. [200]	2020	Deep learning	Unimodal	A motion patch based Siamese convolutional neural network (MSCNN) was developed to address problems with random cropping techniques by extracting the important motion square region
Gowda et al. [45]	2020	Deep learning	Unimodal	proposed the SMART model, which uses temporal segment networks to generate an effective frame selection strategy for video
Ullah et al. [169]	2021	Deep learning	Multimodal	A Multi-view action recognition system utilizes frame level features and Conflux LSTM for accurate recognition and temporal relationships
Streaming				
Soomro et al. [158]	2016	Handcrafted features	Unimodal	Super-pixels are transformed into frames, extracted action segments, and SVM scores are dynamically programmed for action forecasting, with pose and appearance data available online
Jalal et al. [65]	2017	Handcrafted features	Multimodal	Study uses spatiotemporal multi-fused features for accurate online activity recognition, capturing spatial and temporal information for real-time body movements recognition
Zolfaghari et al. [209]	Deep learning	Unimodal		Proposed ECO uses feature representation and CNN network to reduce overhead, extracting complex information from still pictures for ac- curate predictions
Lin et al. [94]	2019	Deep learning	Multimodal	TSM module enhances 3D CNN performance with unidirectional and bidirectional features, focusing on real-time processing and accuracy
Yang et al. [190]	2019	Deep learning	Multimodal	The paper introduces an incremental adaptive deep model (IADM) for real-world incremental data scenarios, addresses capacity scalability and sustainability challenges, and out performs state-of-the-art methods in incremental learning scenarios

Table 9 (continued)

Ref	Year	Approach	Modality	Description
Reinolds et al. [136]	2022	Deep learning	Multimodal	The authors found visual cues in videos provided more informative and discriminative features while combining video and audio inputs improved performance
Ullat et al. [168]	2021	Deep learning	Unimodal	A new method for sequential extraction uses a CNN model and deep skip-gated recurrent unit to learn patterns

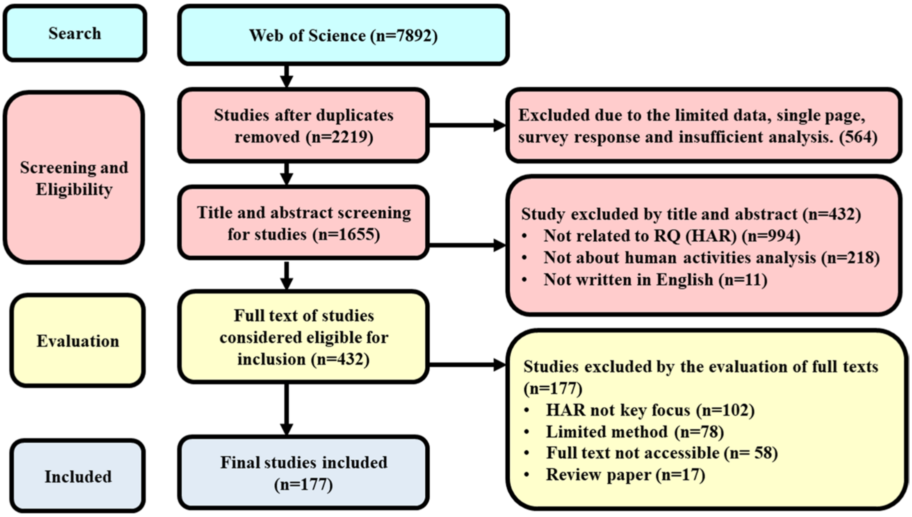

Fig. 1 PRISMA flow diagram used for article selection

4.3 User Activities

User activities are categorized as single activities and group activities, focusing on personal growth and well-being.

4.3.1 Single HAR Datasets

Single human activity datasets may include MuHAVi [154], IXMAS [177] , Weiz- man [107] and KTH [146] etc. It encompasses a wide range of scenarios, such as daily routines, social interactions, and professional engagements. Human behavior is complex and varies in motion as well as appearance. Hsu et al. [60] used unsupervised learning to label video segments for psychiatric patients using N-cut, SVM, and CRF methods, revealing complex human behavior. This approach allowed them to analyze and classify the different behaviors exhibited by the patients, providing valuable insights into their mental health. Combining these techniques enabled a more comprehensive understanding of human behavior and its potential applications in psychiatric care. The KTH dataset was created by Sweden's Royal Institute of Technology [146] in 2004. The dataset contains 2391 actions across four contexts. It consists of 25 sets with six different human activities performed up to five times by 25 participants. The videos have segments that last 4 seconds on average and are shot with a single camera against a static background. MuHAVi [154] was developed in 2010 at Kingston University. It focuses on silhouette-based methods for identifying human activity. The 14 actors repeated the action scenes in the videos 14 times. Eight randomly placed, non-synchronized cameras were used for this, one on each platform's four corners and four sides.

The proposed framework by Ko and Sim [82] utilises the power of deep convolutional networks to identify abnormal human behaviour in RGB images effectively. By incorporating three modules, it successfully addresses the challenges of separating object entities,

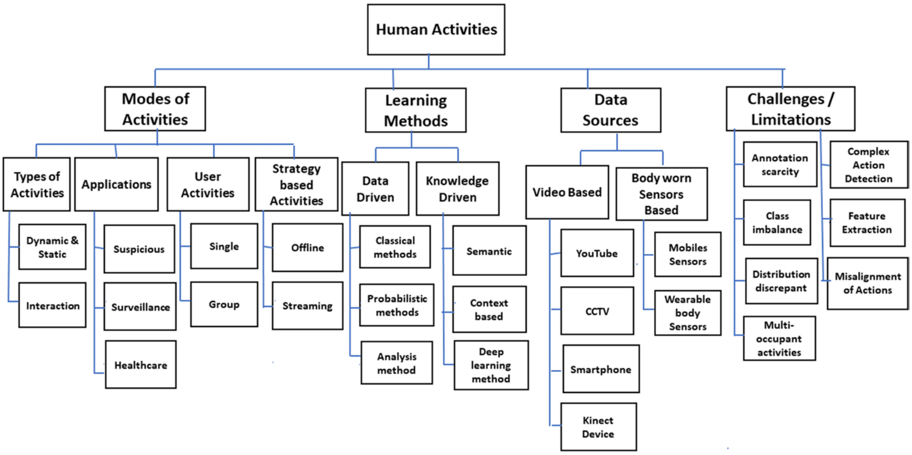

Fig. 2 Exploring the HAR Taxonomy: A Multidisciplinary Framework for Classifying Human Activities

extracting posture features, and detecting abnormal behaviour using LSTM. This unified approach showcases promising results in real-world surveillance scenarios, making it a valuable tool for smart surveillance systems. Overall, the algorithm presented by J. Zhang et al. [198] demonstrates promising results in improving the detection of abnormal behavior in narrow-area scenes captured by CCTV cameras. The adaptive transformation mechanism allows the algorithm to adjust its parameters according to the specific characteristics of the scene, resulting in better detection of abnormal behaviors. Additionally, the improved pyramid L-K optical flow method enhances the algorithm's ability to track and analyze motion patterns, further enhancing the accuracy of abnormal behavior detection.

TransTM is designed to overcome the limitations of existing methods by eliminating the need for complex data cleaning and extending recognition capabilities to include single-person activities and human-to-human interactions. By leveraging the power of the Multiscale Transformer, TransTM can capture nuanced behavioural features with higher accuracy and efficiency. This novel approach offers advantages over traditional CNN and LSTM-based methods, providing enhanced data fitting power, improved generalisation, and greater scalability. With these advancements, TransTM holds great promise for applications in various domains, such as healthcare monitoring, smart homes, and security systems [98].

4.3.2 Group HAR datasets

To detect group activity, the GLIL architecture allows for capturing both local and global dependencies within the group, enhancing the accuracy of activity detection. By incorporating P-LSTM and GLSTM, Shu et al. [151] were able to effectively model the interactions between individuals at both a micro and macro level, resulting in improved performance in group activity recognition tasks. P- LSTM is combined with G-LSTM to learn person-level residual features, tested on CAD and VD datasets for advanced results compared to other methods. The results showed that combining P-LSTM and GLSTM outperformed other methods in accurately predicting collective activities. This suggests that

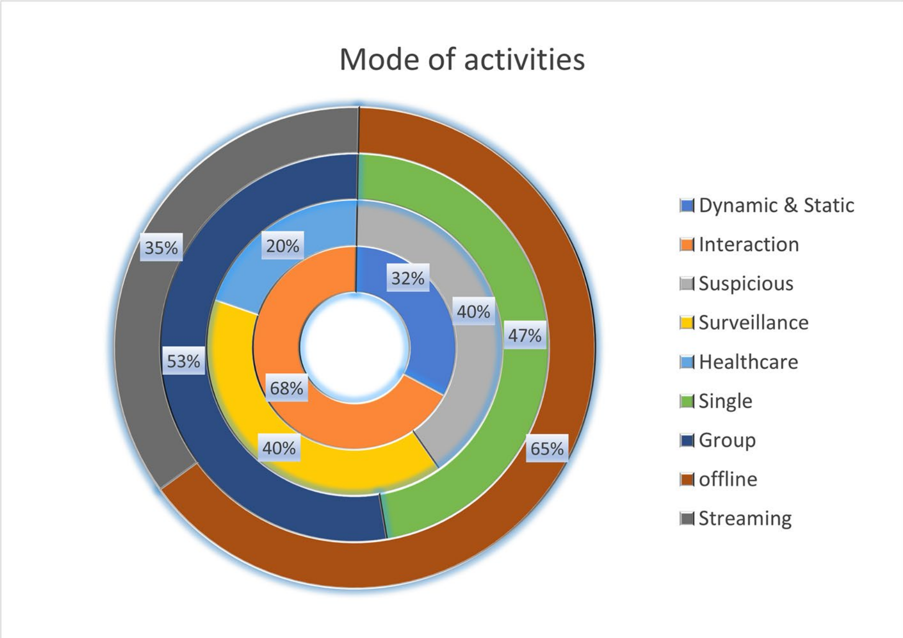

Fig. 3 The frequency of activity mode used in current HAR literature

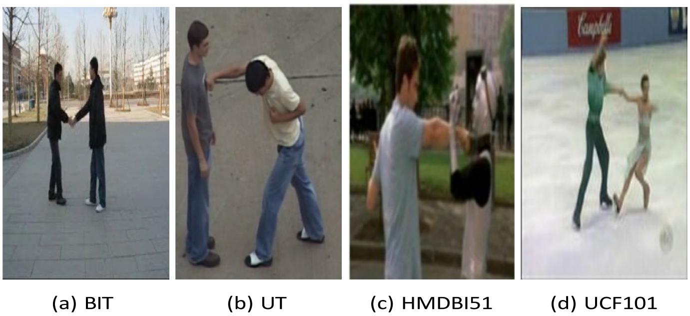

Fig. 4 Example frames from the human-to-human interaction datasets. (**a**) handshake (**b**) punch (**c**) dance

considering local and global interactions between people can significantly improve the performance of activity recognition models.

Group activities often involve multiple people engaging in a shared task or experience, fostering collaboration and social interaction. However, due to their complexity and the potential for variations in execution, it can be challenging to accurately track or distinguish individual contributions within these activities. Researchers [22] propose methodologies

(a) UCF Sports (b) Olympic-sport (c) UCF50 (d) VIRAT

Fig. 5 Example frames from the human-to-object interaction datasets. (**a**) and (**b**) weight lifting (**c**) violin (**d**) GOV (Getting out of a vehicle)

(a) MuHAVi (b) Weizman (c) KTH (d) IXMAX

Fig. 6 Example frames from the Single human action datasets (a)(b) and (c) are walking, (d) hand waving

for the recognition of complicated activities, which facilitate the identification of declared activities. These methods utilize advanced technologies such as computer vision and machine learning algorithms to analyze the movements and interactions of individuals in group activities. By capturing and analyzing data such as body poses, facial expressions, and spatial relationships, these methods can accurately identify and attribute individual contributions within complex activities. This helps in understanding the dynamics of group interactions and enables effective evaluation and feedback for each participant's performance.

Human activities like parties and weddings occur in a certain setting, as do high-level activities that reflect how people interact [63]. These high-level activities include diplomatic negotiations, business meetings, and academic conferences. The context in which these activities occur can greatly influence the outcomes and dynamics of the interactions. Understanding and navigating the specific context is crucial for successful engagement and effective communication among individuals involved in these activities. Figure 7 shows some group dataset frames.

(a) MuHAVi (b) IXMAX (c) Weizman (d) KTH

Fig. 7 Example frames from the group human action datasets. (a) boxing, (b) football, (c) kick, (robbery)

4.4 Strategy-based activities

When working with real-time systems like surveillance or monitoring applications, HAR can be done both online (via a live stream) and offline (via stored videos).

4.4.1 Offline

Offline HAR analyses pre-recorded videos, allowing for more extensive processing and analysis. On the other hand, online HAR requires real-time processing of a live stream, enabling immediate response and decision-making. Modality source refers to unimodal or multimodal methods requiring the input of single or multiple modality inputs [92]. Multimodal methods have the advantage of capturing more comprehensive and nuanced information, leading to improved accuracy in activity recognition. Additionally, combining different modalities can provide a more robust and reliable system by compensating for limitations or uncertainties in individual modalities[189].

Simple offline-unimodal methods [57, 58, 96, 97, 121, 141, 176] are included in HAR, as are complex online-multimodal systems [94] that use real-time data from multiple sources such as sensors, cameras, and microphones. These systems can accurately recognize and classify activities in real-time, making them suitable for applications such as healthcare monitoring and smart home automation. HAR systems use handcrafted feature-based or learning-based approaches, analyzing data from a single modality for activity recognition. On the other hand, online multimodal systems integrate data from multiple modalities in real-time, such as combining accelerometer and gyroscope data for more accurate activity recognition. These advanced systems take advantage of the complementary nature of different modalities to enhance the overall performance of activity recognition.

4.4.2 Streaming

The live stream typically consists of video data captured by cameras or sensors that capture human movements and actions in real-time. The HAR model then processes this data, which analyzes and classifies the activities being performed, enabling AR/VR applications to respond accordingly or self-driving cars to make informed decisions based on the detected activities. Most methods are for offline systems, not real-time security surveillance systems. However, recent advancements in deep learning techniques have enabled the development of real-time online streaming HAR models. These models can process video frames in real-time, making them ideal for applications that require immediate activity recognition and response, such as security surveillance systems.

Jalal et al. [65]used depth differential silhouettes (DDS) and human temporal points to identify online activity, considering skeletal joint characteristics. They used code vectors to reduce computational complexity and used a machine learning algorithm to classify online activities based on extracted features. This approach allowed for real-time monitoring and identification of specific online behaviors, enabling a better understanding of user engagement and interaction patterns.

Zolfaghari et al. [209] developed a lightweight algorithm for real-time activity recognition using HMM and depth maps. This approach suits resource-constrained environments and minimizes the data required for accurate predictions. Combining the 3D and 2D networks in the ECO architecture allows for a more comprehensive understanding of human activity by leveraging temporal and spatial information. Collecting frames from both the current sequence and the following series makes the predictions generated more accurate and efficient, reducing computational complexity and minimizing data overhead. This approach enhances the real-time online activity identification process.

The temporal Recurrent Network (TRN) model considers the sequential nature of actions by incorporating the temporal dependencies between frames. By considering past and future actions, the TRN model aims to improve the accuracy of activity predictions. Additionally, Xu et al. [183] demonstrated the effectiveness of their approach through extensive experiments on various datasets, achieving state-of-the-art results in online HAR.

The temporal shift module introduced by Lin et al. [94] allows for improved accuracy in both online and offline recognition tasks. By considering upcoming video frames in online recognition, the model can make predictions based on the most recent information available, while bidirectional offline recognition considers both past and future frames to enhance accuracy further. This approach has been validated through experiments conducted by Lin et al., showcasing its effectiveness in various datasets and achieving state-of-the-art results in HAR. Gao et al. [41] suggested a weakly-supervised online action detection system to enhance online action detection from untrimmed movies. The offline temporal proposal generator (TPG) analyzes video frames and creates labels based on temporal patterns, providing an initial understanding of video activities. The online action recognizer (OAR) refines this understanding by continuously analyzing the video stream in real-time, detecting specific actions. This combination of offline and online processing ensures accurate and efficient activity detection in both trimmed and untrimmed videos, making it suitable for real-time applications. However, real-time settings require rapid identification based on fresh frames, making online recognition more sophisticated. The WOAD architecture simplifies decision-making in offline situations, making it more efficient in real-time scenarios.

5 Learning Methods

HAR methods are categorized into knowledge-driven and data-driven approaches [14, 182, 203]. Data-driven methods use labeled training data, machine learning algorithms, and deep neural networks for pattern identification and relationships. Knowledge-driven methods rely on prior knowledge or expert-defined rules, often using domain-specific knowledge or predefined models. Both approaches have strengths and limitations. In Figure 8, the classification of HAR learning methods is presented.

5.1 Data Driven

Data-driven learning methods include classical, probabilistic, and analysis Methods, focusing on traditional statistical techniques, modelling uncertainty, and ex- ploring data for insights and decision-making.

5.1.1 Classical methods

SVM is a supervised classifier that learns quickly and efficiently, making it ideal for data analysis and classification. It is used for binary classifiers and can handle high-dimensional data efficiently using the kernel trick. SVM's versatility in handling non-linear decision boundaries makes it suitable for various classification problems in various domains [124, 142, 180, 204]. Bustoni et al. [25] aims to evaluate and compare the performance of three machine learning techniques, namely SVM, K-Nearest Neighbors (K-NN), and Random Forest, in the classification of sensor data related to human motion activities. The evaluation criteria include accuracy, precision, recall, and computational speed, to identify the most effective method among the three. This study will employ the SVM method to incorporate stochastic gradient descent and a support vector classifier utilizing a radial basis function (RBF) kernel. Noori et al. [117] proposed method demonstrates a 92.4% accuracy

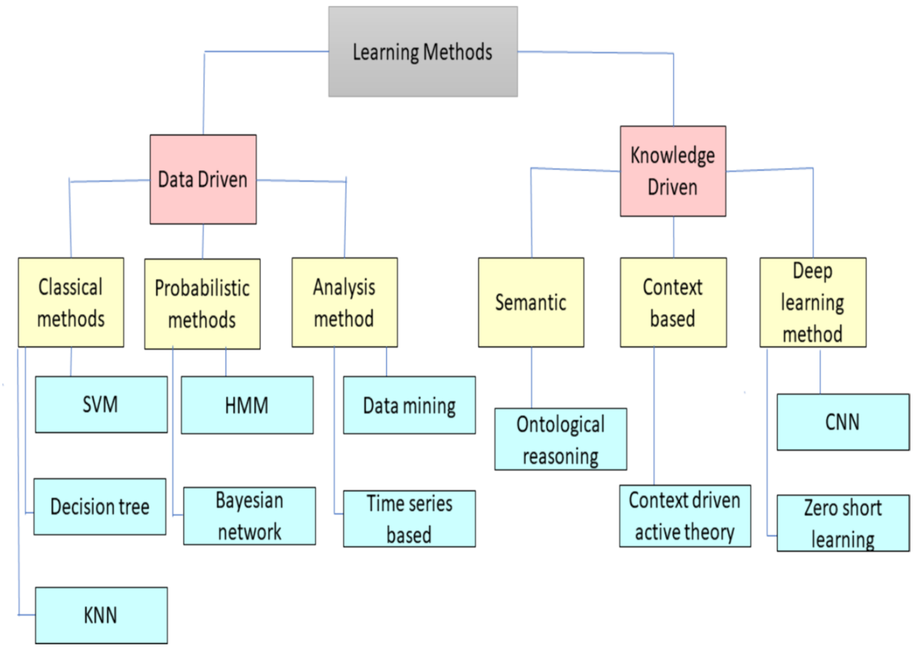

Fig. 8 Classification of HAR learning methods

rate when applied to a publicly accessible activity dataset, surpassing the performance of traditional techniques such as support vector machines and decision trees. The implementation of this activity recognition system holds potential for advancing research in the fields of image processing and computer vision.

5.1.2 Probabilistic methods

Probabilistic learning methods for data-driven learning include HMM and Bayesian networks. These methods are widely used in various fields, such as speech recognition, natural language processing, and computer vision. They are effective at modelling complex patterns and capturing dependencies between variables. The approach employed a combination of the bayes classifier and convolutional neural network. The input for the analysis consists of the KTH dataset, which is utilized to detect moving human targets using a Kalman filter. The Kalman filter algorithm is employed to extract various features from the detected targets, including the length-width ratio, entropy, and Hu invariant moment. A CNN is trained using the KTH dataset, resulting in enhanced accuracy for the detection of abnormal activities such as falling, fighting, and stumbling. This integration facilitates the robust and precise identification of both typical and atypical human activities in real-time situations [95].

The unknown micro-movements are classified using a combination of SVM and hidden markov models (HMM) in the procedure. The findings indicate that when 80% of the known labels are utilized, the process yields results comparable to those achieved in supervised paradigms. This approach mitigates the need for labelling many examples and minimizes the associated economic expenses, thereby addressing the limitations inherent in machine learning algorithms. Using the SVM and Markov Hidden Model, they could identify micro-movements and detect human physical activity accurately. This approach achieved comparable results to supervised paradigms, significantly reduced the need for labelled examples and minimized the associated economic costs [106].

5.1.3 Analysis methods

Belhadi et al. [21] emphasize the need for diverse data mining methodologies and deep learning structural designs to improve outcomes. They suggest that exploring new approaches and techniques could lead to significant advancements in data mining and deep learning. The algorithm was divided into two categories. The first uses data mining and knowledge discovery to detect collective abnormal actions, while the second uses deep CNN to detect collective abnormal behavior. These techniques were chosen to improve the accuracy and efficiency of the algorithm in identifying and analyzing pedestrian behavior in smart cities. By combining data mining, knowledge discovery, and deep CNN, the algorithm can effectively detect both individual and collective abnormal behavior patterns.

Rueda et al. [109] developed a deep neural network architecture to distinguish static and dynamic activities using multichannel time-series data from body-worn devices. The most favorable outcomes were achieved using the Opportunity platform and an industrial dataset. This limitation can be addressed using the proposed deep neural network, as it can accurately classify static and dynamic activities without relying on predefined time windows. By analyzing the multichannel time-series signals, the network can detect the start and end of activities in real- time, making it a valuable tool for applications such as activity recognition in healthcare and sports monitoring. Furthermore, the success of this approach on Opportunity and the industrial dataset demonstrates its potential for generaliz- ability across different domains and sensor configurations. Moukafih et al. [108] The LSTM-FCN model is proposed to identify instances of aggressive driving through time series classification, aiming to address this issue. LSTM can capture temporal dependencies in the driving data, while FCN can efficiently extract features from the input. By treating aggressive driving detection as a time series classification task, this proposed model aims to accurately identify and classify instances of aggressive driving behavior.

5.2 Knowledge Driven

Knowledge learning methods include semantic, context-based, and deep learning methods. Semantic learning methods focus on understanding the meaning and relationships between words and concepts. Context-based learning methods con- sider the surrounding context to enhance understanding and interpretation. Deep learning methods involve training neural networks with multiple layers to extract complex patterns and make accurate predictions. These different approaches offer diverse strategies for acquiring knowledge in various domains and have proven to be effective in different applications such as natural language processing, computer vision, and speech recognition.

5.2.1 Semantic

Semantic learning uses SVMs to classify human activities based on semantic representations. This method has proven effective in various domains, such as healthcare monitoring, video surveillance, and gesture recognition. By utilizing SVMs and semantic representations, it contributes to the advancement of intelligent systems and real-time activity recognition. Recently, using knowledge-driven approaches like ontologies to make semantic smart homes has gotten a lot of attention because of their flexibility, reasoning, and ability to represent knowledge. Ontologies provide a structured and formal way to represent and organize knowledge, allowing smart homes to understand and reason about the context and meaning of various devices, actions, and events within the home environment. By incorporating ontologies into smart homes, it becomes possible to create intelligent systems that can adapt to user preferences, anticipate needs, and automate tasks based on a deeper understanding of the underlying semantics [210].

5.2.2 Context based

Vernikos et al presents a method for HAR using handcrafted features from 3D skeletal data and contextual features learned by a trained deep CNN. The approach improves recognition accuracy in arm gesture recognition by combining contextual features with handcrafted features. The handcrafted features from the 3D skeletal data give information about the spatial relationships and joint angles, while the contextual features learned by the deep CNN capture higher-level patterns and context in the arm gestures. Combining these two types of features achieves a more comprehensive representation of the arm gestures, leading to improved recognition accuracy [171, 174, 201].

5.2.3 Deep learning method

CNNs are widely acknowledged as a leading deep learning technology, with numerous reliably fitted layers. It has been demonstrated to be highly accurate and is widely used in various computer vision tasks. CNNs are especially excellent at im- age classification, object detection, and image segmentation. The ability of CNNs to automatically learn hierarchical features from raw data makes them well-suited for handling complex visual patterns and achieving state-of-the-art performance. Another approach, proposed in [88], involves using a graph convolutional network to directly process the skeletal data without converting it into visual representations. This method has shown promising results in capturing temporal dependencies and achieving state-of-the-art performance in action recognition tasks.

Zero-Shot Learning (ZSL) addresses the issue of large annotated data in supervised action recognition. Two approaches are proposed an inverse autoregressive flow-based generative model and a bi-directional adversarial GAN. The proposed approach uses unlabeled data from unseen classes to train the model. Zero-shot learning is a new technique that has caught the attention of researchers. It can automatically recognize actions from new or unseen classes without the need for explicit training in those classes. This is particularly useful when collecting annotated data for every possible action class is impractical or time-consuming. By leveraging unlabeled data, ZSL allows for more flexible and scalable action recognition systems [104].

The best-performing methods STF+LSTM [139], SAM-SLR [67] Ensemble- NTIS [59], MViT-SLR [118] have been widely recognized in the field of machine learning and have shown promising results in various applications. These Methods have been extensively studied and compared against other state-of-the-art approaches, consistently outperforming them in terms of accuracy and efficiency. Researchers and practitioners alike have adopted these methods due to their robustness and ability to handle complex datasets with high-dimensional features.

Transformer models are currently catching on among deep learning researchers, mainly due to their ability to efficaciously detect long-range dependencies and process sequential data, including natural language, with more precision compared to the previously prevailing practice. At first, instead of natural language processing, transformers were adopted for computer vision and speech recognition, which are now used in other fields. In contrast with the recurrent neural network (RNNs) architecture, which is based mostly on self-attention mechanisms, the transformers pay attention to the importance of different input elements. Thus, they can capture wide dependencies from the input sequence without sequential input processing. That is a great feature of transformers that makes them so powerful for big projects, as they enable fast training and inference speed. Also, transformers have completed the initial steps in pre-training and fine-tuning. In this process, models get all the huge untagged data and then fine-tune it for a particular task. The approach has been able to produce excellent results. Such Transformer models have shown their ability to mimic interrelations of dependency and their wide applicability in various domains. These serve as a great foundation for modern deep learning research and its applications.

The data from a single-factor ANOVA comparison of two groups, data-driven and knowledge-driven, is shown in the table. The count, sum, average, and variation for each group are detailed in the ”Summary” section. The data-driven group’s statistics are as follows: 3 counts, 8 totals, a 2.668 average, and a 0.334 variance. Similar results can be seen for the knowledge-driven group, which has a count of 3, a sum of 4, an average of 1.34, and a variance of 0.334. The ”ANOVA” part includes the study’s p-value, below the accepted significance threshold of 0.05, of 0.0475, indicating a significant difference in averages between the two groups. There is a considerable difference between the groups, as shown by the 8 F-ratio. The crucial F-value is 7.709 as shown in Table 10, 11

Table 10 Summary of single-factor for learning methods

Groups	Count	Sum	Average	Variance
Data Driven	3	8	2.668	0.334
Knowledge Driven	3	4	1.334	0.334

Table 11 One-way ANOVA

Source-of-Variation	SS	df	MS	F	P-value	F-crit
Between Groups	2.668	1	2.668	8	0.0475	7.709
Within Groups	2.3333	24	0.3663			
Total	4	5				

The distribution of learning methods in the existing literature on HAR suggests a significant shift towards independent variable data-driven approaches, with classical methods accounting for 38% and probabilistic methods for 25%. Additionally, analysis methods make up 37% of the studied approaches Table 12. In contrast, independent variable knowledge-driven methods exhibit a more balanced distribution, with semantic and context-based methods each accounting for 25%, while deep learning methods hold the majority share of 50% as shown in Figure 9.

6 Data sources

HAR involves detecting and classifying human actions and behaviors using techniques such as machine learning [93] and deep learning [131, 132]. It plays a crucial role in improving the accuracy and efficiency of automated systems by enabling them to understand and respond to human activities in real-time. The nature of the data generated by various sources, including but not limited to videos, photos, or signals, significantly influences the methodologies employed in HAR. Video in HAR is crucial for security, surveillance, and detecting human actions. Vision- based HAR uses video sources like CCTV and smartphones to identify and forecast activities. Sensor-based HAR is promising for elderly individuals, analyzing sensor data from mobile phones and body-worn sensors. These sensors encompass a range of technologies, including gyroscopes, accelerometers, Bluetooth, and sound sensors, among others. HAR has garnered significant attention in computer vision due to its widespread application across various domains, including healthcare, HCI, security, and surveillance. The utilization of video in HAR holds significance due to its applications in security, surveillance, and the identification and analysis of human activities and behaviors. Vision-based HAR has been widely employed in academic research, utilizing diverse video sources such as closed-circuit television (CCTV), smartphone cameras, Kinect devices, and social media platforms like YouTube. Its primary objective is to identify and anticipate activities in video streams. On the other hand, sensor-based HAR has emerged as a highly promising assistive technology, particularly for supporting elderly individuals in their day- to-day activities. The study centres on the analysis of sensor data obtained from various sources, including mobile phone sensors and body wearable sensors such as gyroscopes, accelerometers, Bluetooth, and sound sensors, among others.

Table 12 Learning methods used in the field of literature of HAR

Ref	HMM	DT	KNN	SVM	BN	Data mining	RF	LSTM	CDA	CNNZSL		PCA
Bustoni et al.[25]	—	—	✓	✓	—	—	—	—	—	—	—	—
Haojie et al.[100]	—	—	—	✓	—	—	✓	✓	—	✓	—	—
Noori et al. [117]	—	✓	—	✓	—	—	—	—	—	—	—	—
Sanal Kumar et al. [144]	—	—	✓	✓	—	—	—	—	—	—	—	—
Liu et al. [95]	—	—	—	—	✓	—	—	—	—	✓	—	—
Morales Garc´ıa et al. [106]	✓	—	—	✓	—	—	—	—	—	—	—	—
Belhadi et al. [21]	—	—	—	—	—	✓	—	—	—	✓	—	—
Qin et al. [130]	✓	—	—	✓	—	—	—	—	✓	—	—	—
Ebrahimpour et al. [38]	—	—	—	✓	—	—	—	—	✓	✓	—	—
Vernikos et al. [171]	—	—	—	✓	—	—	—	—	✓	✓	—	—
Sukor et al. [161]	—	✓	—	✓	—	—	—	—	—	✓	—	✓
Mohan et al. [105]	—	—	—	✓	—	—	—	—	—	✓	—	✓

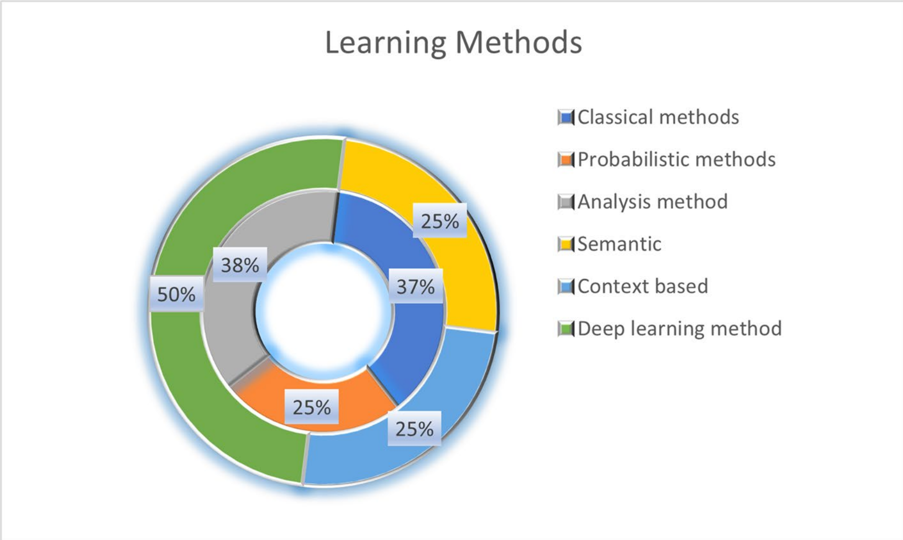

Fig. 9 The frequency of Learning method used in current HAR literature

Data sources are divided into two categories: vision based and sensor-based. Vision-based data sources rely on visual information captured by cameras or other imaging devices. These sources provide data through image processing and analysis techniques, allowing for object detection, recognition, and tracking. On the other hand, sensor-based data sources utilize various sensors such as accelerometers, gyroscopes, and GPS receivers to collect information about physical parameters like motion, orientation, and location. These sources offer valuable insights into environmental conditions and enable the measurement of real-time changes in the surroundings.

6.1 Video-based sensors

Security is a major issue in human society, prompting the construction of homes as well as investments in front-door locks and CCTV systems [15, 129, 187]. Luiz Paulo Oliveira Paula et al. [123] present a unique algorithm for front door security (FDS) that employs HAR to identify four security concerns at the front door with an accuracy rate of 73.18%. The algorithm detects and categorizes actions from CCTV cameras by combining GoogLeNet and BiLSTM hybrid networks. It can also detect door tampering by using powerful motions such as kicking, punching, or hitting. The FDS algorithm detects gun violence at the entrance, further strengthening security.

The YouTube dataset for HAR is a comprehensive collection of videos labelled and annotated to identify various human activities. It includes various activities such as walking, running, dancing, cooking, and playing sports. This dataset is a valuable resource for researchers and developers working on machine learning algorithms and computer vision techniques to accurately recognize and classify human activities in videos [1, 3].

Syed K. Bashar et al. [18] propose a neural network model for smartphone- based HAR using activity-driven, manually created features. The model uses a neighborhood component analysis-derived feature selection method to select significant features from temporal and frequency domain parameters. A deep neural network with four hidden layers classifies input information into multiple categories. The model outperforms state-of-the-art methods while utilizing fewer features, highlighting the importance of careful feature selection in HAR.

The Kinect device dataset is a valuable resource for HAR. It provides a wide range of motion and depth data, allowing for accurate analysis and understanding of human movements. This dataset has been widely used in research and development of activity recognition algorithms, contributing to advancements in fields such as healthcare[54, 62], gaming, and robotics [8, 93, 145].

6.2 Body-worn-based sensors

The mobile sensor dataset for HAR is a collection of data obtained from various sensors embedded in mobile devices. These sensors include accelerometers, gyroscopes, magnetometers, and GPS receivers, among others. The dataset analyses and recognises different human activities such as walking, running, sitting, and standing. By studying the patterns and characteristics of sensor data during these activities, machine learning algorithms can be trained to accurately classify and predict human activities based on real-time sensor readings. This dataset plays a crucial role in developing applications related to health [89, 90, 119, 159, 197].

Sensor-based HAR comprises five steps: sensor selection, data collection, feature extraction, model training, and model testing. The wearable body sensor dataset for HAR is a comprehensive collection of data gathered from various sensors placed on the human body. This dataset provides valuable insights into the movements and actions of individuals, allowing for accurate recognition and analysis of different activities. The data includes information such as accelerometer readings, heart rate measurements, and GPS coordinates, enabling researchers to develop advanced algorithms and models for activity recognition systems [188]. A novel deep neural network was proposed by Rueda et al. [109] for identifying static and dynamic activities from multichannel time-series signals collected from a variety of wearable devices. Alghyaline [6] proposed a method for detecting static and dynamic activities in over 32 fps CCTV camera videos in real time using YOLO object detection, Kalman filtering, and homography. The accuracy of the BEHAVE dataset was found to be 96.9%, while that of the CCTV datasets was only 88.4%. D'Arco et al. [37]The proposed system uses inertial and pressure sensors to identify daily activities, achieving 94.66% accuracy when used in tandem. Inertial sensors capture motion better, while pressure sensors capture stillness.

The data from a single-factor ANOVA comparison of two groups video-based and body-worn sensors-based, is shown in the table 13, 14. The "Summary" section includes information on each group's count, sum, average, and variation. The video-based group's statistics are as follows: 4 counts, 4 totals, a 1 average, and a 0 variance Similar results can be seen for the body worn sensors-based group. The given data in the tables indicates that there are 2 counts with a sum of 15. These numbers have an average of 7.5. Furthermore, the variance of 12.5 suggests that the numbers have a significant spread from their average value, indicating a considerable difference between them. The "ANOVA" part includes the sources of variation: SS, df, MS, F, p-value, and critical F-value. The study's p-value, below the accepted significance threshold of 0.05, of 0.0132 indicates a significant

Table 13 Summary of single-factor for data sources

Groups	Count	Sum	Average	Variance
Video Based	4	4	1	0
Body worn Sensors Based	2	15	7.5	12.5

Table 14 One-way ANOVA

Source-of-Variation	SS	df	MS	F	P-value	F-crit
Between Groups	56.4	1	56.4	18.2	0.013	7.708
Within Groups	12.5	4	3.125			
Total	68.84	5				

difference in averages between the 2 groups. There is a considerable difference between the groups, as shown by the 18.2 F-ratio. The crucial F-value is 7.708 Table 15.

These percentages indicate the distribution of data sources commonly used in HAR research. Video-based sources, such as YouTube, CCTV, smartphones, and Kinect devices, contribute significantly to the existing literature, with each source accounting for 25% of the focus. Additionally, body-worn sensors play a smaller role in HAR studies, with mobile sensors representing 33% and wearable body sensors comprising 67% of the analyzed data sources as shown in Fig. 10.

Table 15 Data Sources used in literature

	Video Based				Body	-Worn Sensors
Ref	YouTube	CCTV	Smartphones	Kinect-Devices	Mobile-Sensors	Wearable-Sensors
Bashar et al. [18]	—	—	✓	—	—	—
Paula et al. [123]	—	✓	—	—	—	—
Abadi et al. [1]	✓	—	—	—	—	—
Sawanglok et al.[145]	—	—	—	✓	—	—
Lazaridis et al. [87]	✓	✓	—	—	—	—
Sorkun et al. [159]	—	—	—	—	✓	—
Kang et al. [69]	✓	✓	✓	—	—	—
Guo et al. [49]	—	—	—	—	✓	✓
Gupta et al. [52]	✓	✓	—	—	—	—
Yang et al. [188]	—	—	—	—	—	✓
Khan et al. [75]	✓	✓	—	—	—	—
Moukafih et al. [108]	—	—	✓	—	✓	—

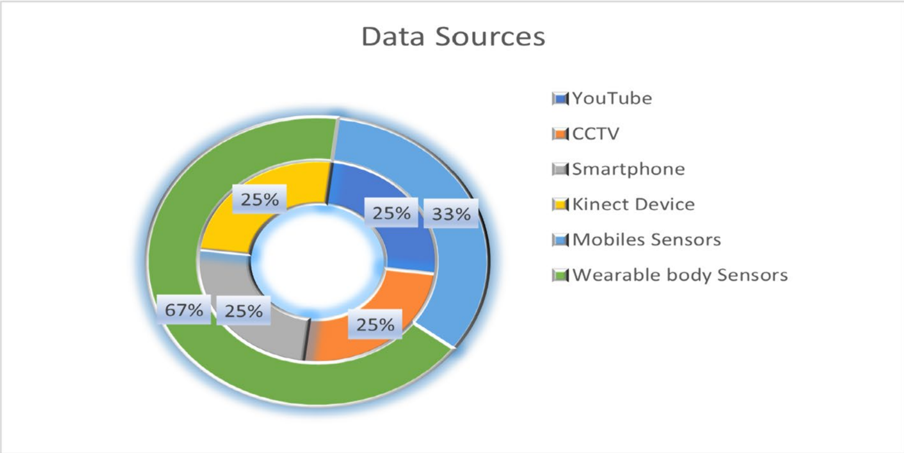

Fig. 10 The frequency of data sources used in current HAR literature

7 Video action recognition for training methodologies

The recognition of the video actions encompasses the task of not only identifying but also classifying actions and activities on recorded video sequences. In addition to the training platforms, many approaches aim to improve the precision and reliability of video action identification models. Three prominent methodologies are supervised, Sami supervised, and unsupervised learning.

7.1 Supervised learning

A supervised learning methodology is widely used for video action recognition training. It requires labeled training data, which is made up of video clips with the corresponding action label. Supervised learning demands a lot of data, which has been labeled with accuracy, to cover different action categories. Yet, getting big enough video datasets classified could be tedious and costly. Through models like the convolutional neural network, the network learns to associate specific visual cues and patterns with the action labels, and these are achieved by optimization techniques such as backpropagation and gradient descent. Despite that, supervised learning continues to be one of the popular techniques for video action recognition because of its ability to capture discriminating features and achieve high classification accuracy [74].

7.2 Self-Supervised Learning

Automatic teachings is the latest method for learning videos under the idea of recognizable something without asking any human being to annotate it. Rather it uses an intrinsic property of the video data, like its structure or order of the data itself, and plans to stitch different masks to generate the pretext task. This learning mechanism lets the network figure out which features are meaningful so that it can capture the semantics of the body's actions highlighted through these tasks. The temporal order of frames and even the rotation of video patches can also be predicted by pretext tasks in self-supervised learning, including predicting objects' movement direction or even generating videos that only give partial information. After the pre-trained model is treated with the pretext tasks, the features learned can be transplanted into action recognition tasks. Consequently, self-supervised learning is more beneficial when data is unavailable or annotating is expensive [181].

7.3 Semi-Supervised Learning

Semi-supervised learning is the training approach that utilizes datasets consisting of labeled and unlabeled data to better the video action recognition system. In this way, however, the portion of the video dataset is the manual annotation, but the majority is the percentage of the non-annotated video data. Through the employed lever of unlabeled data, semi-supervised learning is factored in to improve the model's generalization and effectiveness. A wide range of solutions, e.g., the ones that combine self-training, co-training, and pseudo-labeling, are used to employ the unlabeled data effectively. A semi-supervised learning model is significant when labeled data is either restricted or expensively accumulated. Still, the unlabeled data possess additional information to guide the machine learning model to have a better and more precise discriminative feature [186].

8 Challenges and Limitations

One of the challenges in HAR is the variability and complexity of human movements. Individuals may perform the same activity differently, making classifying and recognising activities difficult. Additionally, occlusions, such as objects obstructing the view or partial visibility of body parts, can further complicate recognition algorithms. These challenges highlight the need for robust and adaptable recognition systems to handle various scenarios and accurately identify human activities in real-world environments.

8.1 Annotation scarcity

The lack of annotations poses a significant challenge for sensor-based activity recognition systems. It refers to the limited availability of labelled data for training these systems, which hinders their accuracy and performance. This scarcity arises due to the need for expert annotators and the time-consuming process of manually labelling large datasets. Additionally, annotation scarcity can also be attributed to the complexity and diversity of human activities, making it difficult to capture all possible variations in the training data [12]. In addition, video-based activity recognition systems face a critical problem, which is annotation scarcity. The limited amount of annotated video datasets is a stumbling block to building and training models that can achieve accurate video-based human activity recognition. This challenge comes about as a result of the manual and time-consuming nature of the annotation of many videos with exact labels. Due to this, the lack of video data rich in annotations limits the training of strong models that can successfully identify and classify human actions from visual sequences. This issue, however, is often observed for video-based activity recognition systems that heavily rely on annotated data for training and evaluation purposes. Through this process, they will appreciate the special features of annotation deficiency in video-based activity recognition and ultimately see the unique difficulties associated with developing and training models [56].

8.2 Class imbalance

Class imbalance is when one class of data significantly outweighs the other in quantity. This can make it difficult for machine learning algorithms to accurately predict and classify the minority class, as they tend to prioritize the majority class due to its higher representation in the data [50, 191, 192]. The class imbalance definition implies the situation, when some classes have many data samples and the rest have a minor number of instances. Consequently, this challenge will cause problems for the accuracy of the activity recognition models as they will be more biased towards the majority classes neglecting the minority ones hence causing the poor operation on the minority classes.

8.3 Distribution discrepant

Machine learning models exhibit distributional discrepancies due to users, time, and sensors, all of which are present. Users contribute to the distribution discrepancy by generating different types of data based on their preferences and behaviors. Time also plays a role, as the data distribution can change over time due to evolving trends and patterns. Additionally, sensors used to collect data may introduce discrepancies if they have varying levels of accuracy or are affected by environmental factors [173]. Distribution discrepant which happens between the training and the trials phases is often cause of data inconsistency in human activity recognition. It is a realization that the features extracted from the data training set are sub- stantially different from the ones faced via practical real-world testing. This type of issue normally weakens the performance of activity recognition models when they are deployed in practical environments.

8.4 Multi-occupant activities

Multi-occupant activities also require effective communication between the occupants due to the complexity of data association. In such activities, multiple individuals may be interacting with various objects or performing different tasks simultaneously, making it challenging to accurately associate the data generated by each occupant with their respective actions or inputs. Additionally, effective communication becomes crucial to coordinate and synchronize the actions of multiple occupants, ensuring smooth collaboration and avoiding conflicts or misunderstandings [114]. Multi-occupant scenarios are situations where multiple people are involved together in the particular activities at the same time, making category recognition and person identification of each individual practically challenging. This problem stems from the fact that the positions, fleeing dynamics, and back- ground obstructions of people are collectively linked and changing.

8.5 Complex action detection

Complex action detection poses challenging problems in the computer vision field. It involves identifying and understanding human actions in videos, which can vary in scale, viewpoint, and appearance. Additionally, the temporal nature of videos adds another layer of complexity, as actions can unfold over time [185]. High-level action recognition means the ability to detect actions that have more than two sub-actions or steps with the knowledge of the meaning. These actions frequently constitute time-dependent and sequential relation exhibits high-unfeasibility of their detection. Cooking, which might be a recipe, putting together furniture, and performing a sports activity are examples of complex skills.

8.6 Feature extraction

Feature extraction is one of the most difficult problems in computer vision. It involves converting raw data, such as images or videos, into meaningful and representative features that can be used for further analysis or classification tasks. This task is challenging due to the high dimensionality and variability of visual data and the need to capture low-level and high-level information [7].

8.7 Misalignment of actions

Misalignment of actions refers to the challenge of accurately aligning different visual elements or objects within an image or across multiple images. This misalignment can occur due to various factors, such as changes in viewpoint, lighting conditions, occlusions, or deformations in the scene. Solving this problem is crucial for tasks like object recognition, tracking, and image registration in computer vision applications [33, 140, 199]. misalignment action means for the same or different data sources to deliver successfully on time and smoothly. This is about temporal dissimilarities and exceptions where the time stamps in the annotations are not consistent with the actual events timing in the dataset. This challenge may arise from the diversity in data collection methods, latency in data recording or discrepancy in the level of human annotations detail.

In conclusion to address the challenges of Human Activity Recognition (HAR) more efficiently, I suggest using and exploring advanced methods of graph neural networks for the whole purpose of capturing time-dependent dependencies and neighborhoods between actions and sub-actions. Moreover, involvement of attention mechanisms and transformer models is very important as it allow the network to capture the long term dependency and make the identification of complex actions more accurate. To add, related to self-supervised learning techniques that uses, e. g. , the contrastive approach or generative models one can study semantic representations, which leads to the decrease of the need for annotated data. Com- bining real-time feedback with online learning techniques actively eases the way for upgradable and efficient systems for action identification that surpass their actual performance. Therefore, when taking these technical considerations into account, researchers as well as practitioners can help in building the HAR and this can lead to faster development of the HAR programs that are accurate and competitive in identifying and understanding human activities.

9 Conclusion

HAR is now playing an important role in a variety of surveillance and monitoring fields. By categorizing the existing state-of-the-art literature, this review provides a comprehensive understanding of how HAR is being applied across various domains. This review aims to classify the literature by analyzing the various modes of activity, such as interaction, dynamic, & static user activities. It highlights the different application areas, such as suspicious behavior, healthcare systems, and surveillance, and allows us to identify the specific contexts in which HAR is proving to be most effective. We can comprehensively understand how HAR is being utilized in surveillance and monitoring. Some of the data sources discussed in this paper included videos and body-worn sensors, which provide valuable insights into human behavior. These sources allow researchers to gather rich and diverse data for analysis. In terms of learning methods, both data-driven and knowledge-driven approaches were explored during the discussion. Data-driven methods rely on large datasets to extract patterns and make predictions, while knowledge-driven methods incorporate existing domain knowledge into the learning process. Lastly, the review paper highlighted some of the open research challenges researchers currently focus on in this field.

Integration of virtual reality (VR) and augmented reality (AR) technologies into HAR systems is one of the emerging trends and potential future directions in HAR. This integration makes the user experience more immersive and interactive by allowing users to manipulate virtual objects in real-time. Another potential future direction is the development of HAR systems that can adapt to the preferences and requirements of individual users, providing individualized assistance and recommendations based on their specific context and objectives. HAR systems should incorporate artificial intelligence and machine learning algorithms. This enables more precise and real-time activity recognition, as well as the ability to adapt and customize the system based on the preferences of each individual user. In addition, advances in sensor technology, such as the development of wearable devices with multiple sensors, enable HAR systems to capture a broader spectrum of human activities and offer more in-depth insights into user behavior.

Acknowledgements Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R410), Princess Nourah bint Abdul- rahman University, Riyadh,Saudi Arabia. The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number “NBU-FFR-2024-231-10”.

Author's contributions Conceptualization, TFNB, HR, MS and AJ; taxonomy, TFNB; investigation, Data analysis AA; resources NAA; writing—original draft preparation, TFNB; writing—review and editing, TFNB, HR and AJ; statistical analysis, Data analysis, Data curation, Formal analysis; and MS Writing, writing review & editing.

Data availability Data supporting this study will be available refers references included in the paper.

Declarations

Conflict of interest The authors declare no confict of interest.

References

1. Abadi MB, Alashti MRS, Holthaus P, Menon C, Amirabdollahian F (2023) Rhm: Robot house multi-view human activity recognition dataset. IARIA, March. https://hdl.handle.net/2299/27046
2. Abu-El-Haija S, Kothari N, Lee J, Natsev P, Toderici G, Varadarajan B, Vijayanarasimhan S (2016) Youtube-8m: A large-scale video classification benchmark. arXiv preprint arXiv:160908675. https://arxiv.org/abs/1609.08675
3. Ahmad F (2022) Deep image retrieval using artificial neural network inter- polation and indexing based on similarity measurement. CAAI Trans Intell Technol 7(2):200–218
4. Ahmed N, Rafiq JI, Islam MR (2020) Enhanced human activity recognition based on smartphone sensor data using hybrid feature selection model. Sensors 20(1):317
5. Alarfaj M, Waheed M, Ghadi YY, al Shloul T, Alsuhibany SA, Jalal A, Park J (2022) An intelligent framework for recognizing social human-object interactions. Comp Mater Cont 73(1). https://doi.org/10.32604/cmc.2022.025671
6. Alghyaline S (2019) A real-time street actions detection. Int J Adv Comp Sci Appl 10(2). https://doi.org/10.14569/IJACSA.2019.0100243
7. Ali HH, Moftah HM, Youssif AA (2018) Depth-based human activity recognition: A comparative perspective study on feature extraction. Future Computing Inform J 3(1):51–67
8. Ali N, Ullah S, Khan D, Rahman H, Alam A (2023) The effect of adaptive aids on different levels of students' performance in a virtual reality chemistry laboratory. Educ Inf Technol 1–20. https://doi.org/10.1007/s10639-023-11897-0
9. Alkhurayyif Y (2023) Users' information security awareness of home closed- circuit television surveillance. J Inf Sec Cybercrimes Res 6(1):12–23
10. Amrutha C, Jyotsna C, Amudha J (2020) Deep learning approach for sus- picious activity detection from surveillance video. In: 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), IEEE, pp 335–339. https://doi.org/10.1109/ICIMIA48430.2020.9074920
11. Azmat U, Alotaibi SS, Mudawi NA, Alabduallah BI, Alonazi M, Jalal A, Park J (2023) An elliptical modeling supported system for human action deep recognition over aerial surveillance. IEEE Access 11:75671–75685. https://doi.org/10.1109/ACCESS.2023.3266774
12. Babangida L, Perumal T, Mustapha N, Yaakob R (2022) Internet of things (iot) based activity recognition strategies in smart homes: A review. IEEE Sens J 22(9):8327–8336
13. Babiker M, Khalifa OO, Htike KK, Hassan A, Zaharadeen M (2017) Au- tomated daily human activity recognition for video surveillance using neu- ral network. In: 2017 IEEE 4th International Conference on Smart In- strumentation, Measurement and Application (ICSIMA), pp 1–5. https://doi.org/10.1109/ICSIMA.2017.8312024
14. Bahadori S, Williams JM, Collard S, Swain I (2023) Can a purposeful walk intervention with a distance goal using an activity monitor improve individuals' daily activity and function post total hip replacement surgery. a randomized pilot trial. Cyborg Bionic Syst 4:69. https://doi.org/10.34133/cbsystems.0069
15. Ban Y, Liu Y, Yin Z, Liu X, Liu M, Yin L, Zheng W (2024) Micro-directional propagation method based on user clustering. Comp Inf 42(6):1445–1470. https://doi.org/10.31577/cai_2023_6_1445
16. Barekatain M, Mart´ı M, Shih HF, Murray S, Nakayama K, Matsuo Y, Prendinger H (2017) Okutama-action: An aerial view video dataset for con- current human action detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 28–35. https://doi.org/10.1109/CVPRW.2017.267
17. Barman N, Zadtootaghaj S, Schmidt S, Martini MG, M¨oller S (2018) Gam- ingvideoset: a dataset for gaming video streaming applications. In: 2018 16th Annual Workshop on Network and Systems Support for Games (NetGames), IEEE, pp 1–6. https://doi.org/10.1109/NetGames.2018.8463362
18. Bashar SK, Al Fahim A, Chon KH (2020) Smartphone based human activity recognition with feature selection and dense neural network. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), pp 5888–5891. https://doi.org/10.1109/EMBC44109.2020.9176239
19. Baumgartl H, Sauter D, Schenk C, Atik C, Buettner R (2021) Vision-based hand gesture recognition for human-computer interaction using mobilenetv2. In: 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), IEEE, pp 1667–1674. https://doi.org/10.1109/COMPSAC51774.2021.00249
20. Beddiar DR, Nini B, Sabokrou M, Hadid A (2020) Vision-based human activity recognition: a survey. Multimed Tools Appl 79(41–42):30509–30555

21. Belhadi A, Djenouri Y, Srivastava G, Djenouri D, Lin JCW, Fortino G (2021) Deep learning for pedestrian collective behavior analysis in smart cities: A model of group trajectory outlier detection. Inf Fusion 65:13–20
22. Bhardwaj R, Singh PK (2016) Analytical review on human activity recognition in video. In: 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence), IEEE, pp 531–536. https://doi.org/10.1109/CONFLUENCE.2016.7508177
23. Grck BP (2021) Peran kamera pengawas closed-circuit television (cctv) dalam kontra terorisme. J Lemhannas RI 9(4):100–116
24. Bukht TFN, Rahman H, Jalal A (2023) A novel framework for human action recognition based on features fusion and decision tree. In: 2023 4th International Conference on Advancements in Computational Sciences (ICACS), pp 1–6. https://doi.org/10.1109/ICACS55311.2023.10089752
25. Bustoni IA, Hidayatulloh I, Ningtyas A, Purwaningsih A, Azhari S (2020) Classification methods performance on human activity recognition. In: Journal of Physics: Conference Series, IOP Publishing, 1:012027. https://doi.org/10.1088/1742-6596%2F1456%2F1%2F012027
26. Caba Heilbron F, Escorcia V, Ghanem B, Carlos Niebles J (2015) Activitynet: A large-scale video benchmark for human activity understanding. In: Proceedings of the ieee conference on computer vision and pattern recognition, pp 961–970. https://doi.org/10.1109/CVPR.2015.7298698
27. Cai L, Yan S, Ouyang C, Zhang T, Zhu J, Chen L, Liu H (2023) Muscle synergies in joystick manipulation. Front Physiol 14. https://doi.org/10.3389/fphys.2023.1282295
28. Chung J, Wuu Ch, Yang Hr, Tai YW, Tang CK (2021) Haa500: Human-centric atomic action dataset with curated videos. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 13465–13474. https://doi.org/10.1109/ICCV48922.2021.01321
29. Dang LM (2020) Kyungbok min, hanxiang wang, md jalil piran, cheol hee lee, and hyeonjoon moon. sensor-based and vision-based human activity recognition: A comprehensive survey. Pattern Reco 108(107561):3. https://doi.org/10.1016/j.patcog.2020.107561
30. Demir U, Rawat YS, Shah M (2021) Tinyvirat: Low-resolution video action recognition. In: 2020 25th international conference on pattern recognition (ICPR), IEEE, pp 7387–7394. https://doi.org/10.1109/ICPR48806.2021.9412541
31. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A large- scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, IEEE, pp 248–255. https://doi.org/10.1109/CVPR.2009.5206848
32. Dhiman C, Vishwakarma DK (2019) A review of state-of-the-art techniques for abnormal human activity recognition. Eng Appl Artif Intell 77:21–45
33. Di Y, Li R, Tian H, Guo J, Shi B, Wang Z, Liu Y (2023) A maneuvering target tracking based on fastimm-extended viterbi algorithm. Neural Comput Appl. https://doi.org/10.1007/s00521-023-09039-1
34. Diba A, Fayyaz M, Sharma V, Paluri M, Gall J, Stiefelhagen R, Van Gool L (2020) Large scale holistic video understanding. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16, Springer, pp 593–610. https://doi.org/10.1007/978-3-030-58558-7_35
35. Dileep AS, S NS, S S, K F, S S (2022) Suspicious human activity recognition using 2d pose estimation and convolutional neural network. In: 2022 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), pp 19–23. https://doi.org/10.1109/WiSPNET54241.2022.9767152
36. Ding Y, Zhang W, Zhou X, Liao Q, Luo Q, Ni LM (2021) Fraudtrip: Taxi fraudulent trip detection from corresponding trajectories. IEEE Int Things J 8(16):12505–12517. https://doi.org/10.1109/JIOT.2020.3019398
37. D'Arco L, Wang H, Zheng H (2022) Assessing impact of sensors and feature selection in smart-insole-based human activity recognition. Methods Protocols 5(3):45
38. Ebrahimpour Z, Wan W, Cervantes O, Luo T, Ullah H (2019) Comparison of main approaches for extracting behavior features from crowd flow analysis. ISPRS Int J Geo Inf 8(10):440
39. Elharrouss O, Almaadeed N, Al-Maadeed S, Bouridane A, Beghdadi A (2021) A combined multiple action recognition and summarization for surveillance video sequences. Appl Intell 51:690–712
40. Franco A, Magnani A, Maio D (2020) A multimodal approach for human activity recognition based on skeleton and rgb data. Pattern Recogn Lett 131:293–299
41. Gao M, Zhou Y, Xu R, Socher R, Xiong C (2021) Woad: Weakly supervised online action detection in untrimmed videos. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp 1915–1923. https://doi.org/10.1109/CVPR46437.2021.00195

42. Gerling K, Mandryk R et al (2014) Custom-designed motion-based games for older adults: a review of literature in human-computer interaction. Gerontechnology 12(2):68–80
43. Ghadi YY, Waheed M, al Shloul T, A Alsuhibany S, Jalal A, Park J (2022) Automated parts-based model for recognizing human–object interactions from aerial imagery with fully convolutional network. Remote Sens 14(6). https://doi.org/10.3390/rs14061492, URL https://www.mdpi.com/2072-4292/14/6/1492
44. Ghayvat H, Awais M, Pandya S, Ren H, Akbarzadeh S, Chandra Mukhopadhyay S, Chen C, Gope P, Chouhan A, Chen W (2019) Smart aging system: uncovering the hidden wellness parameter for well-being monitoring and anomaly detection. Sensors 19(4):766
45. Gowda SN, Rohrbach M, Sevilla-Lara L (2021) Smart frame selection for action recognition. Proceed AAAI Conf Artif Intel 35:1451–1459
46. Gu C, Sun C, Ross DA, Vondrick C, Pantofaru C, Li Y, Vijayanarasimhan S, Toderici G, Ricco S, Sukthankar R et al (2018) Ava: A video dataset of spatio-temporally localized atomic visual actions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 6047–6056. https://doi.org/10.1109/CVPR.2018.00633
47. Gu Y, Hu Z, Zhao Y, Liao J, Zhang W (2024) Mfgtn: A multi-modal fast gated transformer for identifying single trawl marine fishing vessel. Ocean Eng 303:117711. https://doi.org/10.1016/j.oceaneng.2024.117711
48. Gumaei A, Hassan MM, Alelaiwi A, Alsalman H (2019) A hybrid deep learn- ing model for human activity recognition using multimodal body sensing data. IEEE Access 7:99152–99160
49. Guo J, Mu Y, Xiong M, Liu Y, Gu J (2019) Activity feature solving based on tf-idf for activity recognition in smart homes. Complexity 2019:1–10
50. Guo Y, Chu Y, Jiao B, Cheng J, Yu Z, Cui N, Ma L (2021) Evolutionary dual-ensemble class imbalance learning for human activity recognition. IEEE Trans Emerg Top Comput Intell 6(4):728–739
51. Gupta N, Gupta SK, Pathak RK, Jain V, Rashidi P, Suri JS (2022) Human activity recognition in artificial intelligence framework: A narrative review. Artif Intell Rev 55(6):4755–4808
52. Gupta T, Nunavath V, Roy S (2019) Crowdvas-net: A deep-cnn based frame- work to detect abnormal crowd-motion behavior in videos for predicting crowd disaster. In: 2019 IEEE international conference on Systems, Man and Cybernetics (SMC), IEEE, pp 2877–2882. https://doi.org/10.1109/SMC.2019.8914152
53. Hartmann Y, Liu H, Schultz T (2022) Interactive and interpretable online human activity recognition. In: 2022 IEEE International Conference on Perva- sive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops), IEEE, pp 109–111. https://doi.org/10.1109/PerComWorkshops53856.2022.9767207
54. Hassan FS, Gutub A (2022) Improving data hiding within colour images using hue component of hsv colour space. CAAI Trans on Intel Tech 7(1):56–68
55. Helmi AM, Al-qaness MA, Dahou A, Abd Elaziz M (2023) Human activity recognition using marine predators algorithm with deep learning. Futur Gener Comput Syst 142:340–350
56. Hoelzemann A, Bock M, Van Laerhoven K (2024) Evaluation of video-assisted annotation of human imu data across expertise, datasets, and tools. In: 2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops), IEEE, pp 1–6. https://doi.org/10.1109/PerComWorkshops59983.2024.10503292
57. Hou X, Xin L, Fu Y, Na Z, Gao G, Liu Y, Chen T (2023) A self-powered biomimetic mouse whisker sensor (bmws) aiming at terrestrial and space objects perception. Nano Energy 118:109034. https://doi.org/10.1016/j.nanoen.2023.109034
58. Hou X, Zhang L, Su Y, Gao G, Liu Y, Na Z, Chen T (2023) A space crawling robotic bio-paw (scrbp) enabled by triboelectric sensors for surface identifi- cation. Nano Energy 105:108013. https://doi.org/10.1016/j.nanoen.2022.108013
59. Hrúz M, Gruber I, Kanis J, Boháček M, Hlaváč M, Krňoul Z (2022) One model is not enough: Ensembles for isolated sign language recognition. Sensors 22(13):5043
60. Hsu SC, Chuang CH, Huang CL, Teng R, Lin MJ (2018) A video-based ab- normal human behavior detection for psychiatric patient monitoring. In: 2018 International Workshop on Advanced Image Technology (IWAIT), IEEE, pp 1–4. https://doi.org/10.1109/IWAIT.2018.8369749
61. Hu M, Luo M, Huang M, Meng W, Xiong B, Yang X, Sang J (2023) Towards a multimodal human activity dataset for healthcare. Multimedia Syst 29(1):1–13
62. Hussain S, Rahman H, Abdulsaheb GM, Al-Khawaja H, Khalaf OI (2023) A blockchain-based approach for healthcare data interoperability. Int J Adv Soft Comput Appl 15(2). https://www.i-csrs.org/Volumes/ijasca/IJASCA.230720.06.pdf

63. Iglesias PA, Revilla M (2023) Skills, availability, willingness, expected partic- ipation and burden of sharing visual data within the frame of web surveys. Qual Quant pp 1–22. https://doi.org/10.1007/s11135-023-01670-3
64. Ishikawa Y, Kasai S, Aoki Y, Kataoka H (2021) Alleviating over-segmentation errors by detecting action boundaries. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp 2322–2331. https://doi.org/10.1109/WACV48630.2021.00237
65. Jalal A, Kim YH, Kim YJ, Kamal S, Kim D (2017) Robust human activity recognition from depth video using spatiotemporal multi-fused features. Pattern Recogn 61:295–308
66. Jalal A, Khalid N, Kim K (2020) Automatic recognition of human interaction via hybrid descriptors and maximum entropy markov model using depth sensors. Entropy 22(8):817
67. Jiang S, Sun B, Wang L, Bai Y, Li K, Fu Y (2021) Skeleton aware multi-modal sign language recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3413–3423. https://doi.org/10.1109/CVPRW53098.2021.00380
68. Kamthe UM, Patil CG (2018) Suspicious activity recognition in video surveillance system. In: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), pp 1–6. https://doi.org/10.1109/ICCUBEA.2018.8697408
69. Kang JM, Kang SW, Song YJ (2015) Real time surveilance system development for prevention of school violance and sexual abuse. ICIC Express Letters p 1285. https://www.researchgate.net/profile/Shunping-Lin/publication/282179731_Development_and_application_of_gravity_acceleration_measurement_in_running_kinematic_analysis/links/570dc97308ae2b772e432ce0/Developmentand-application-of-gravity-acceleration-measurement-in-running-kinematic-analysis.pdf#page=13
70. Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L (2014) Large-scale video classification with convolutional neural networks. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp 1725–1732. https://doi.org/10.1109/CVPR.2014.223
71. Kellokumpu V, Zhao G, Pietikäinen M (2008) Human activity recognition using a dynamic texture based method. In: BMVC, vol 1, p 2. https://doi.org/10.5244/C.22.88
72. Khairy H (2022) Statistical features versus deep learning representation for suspicious human activity recognition. In: 2022 4th Novel Intelligent and Leading Emerging Sciences Conference (NILES), pp 126–130. https://doi.org/10.1109/NILES56402.2022.9942440
73. Khan D, Alonazi M, Abdelhaq M, Al Mudawi N, Algarni A, Jalal A, Liu H (2024) Robust human locomotion and localization activity recognition over multisensory. Front Physiol 15. https://doi.org/10.3389/fphys.2024.1344887
74. Khan MA, Mittal M, Goyal LM, Roy S (2021) A deep survey on supervised learning based human detection and activity classification methods. Multimed Tools Appl 80(18):27867–27923
75. Khan SD (2019) Congestion detection in pedestrian crowds using oscilla- tion in motion trajectories. Eng Appl Artif Intell 85:429–443
76. Khan ZN, Ahmad J (2021) Attention induced multi-head convolutional neural network for human activity recognition. Appl Soft Comput 110:107671
77. Khodabandelou G, Moon H, Amirat Y, Mohammed S (2023) A fuzzy con- volutional attention-based gru network for human activity recognition. Eng Appl Artif Intell 118:105702
78. Kim K, Jalal A, Mahmood M (2019) Vision-based human activity recognition system using depth silhouettes: A smart home system for monitoring the residents. J Electr Eng Technol 14:2567–2573
79. KL BJ, VV (2023) Deep maxout network for human action and abnormality detection using chronological poor and rich optimization. Comput Methods Biomech Biomed Eng: Imaging Visual 11(3): 758–773. https://doi.org/10.1080/21681163.2022.2111720
80. KL BJ et al (2021) Chronological poor and rich tunicate swarm algorithm integrated deep maxout network for human action and abnormality detection. In: 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT), IEEE, pp 1–9. https://doi.org/10.1109/ICECCT52121.2021.9616637
81. Kliper-Gross O, Hassner T, Wolf L (2011) The action similarity labeling challenge. IEEE Trans Pattern Anal Mach Intell 34(3):615–621
82. Ko KE, Sim KB (2018) Deep convolutional framework for abnormal behavior detection in a smart surveillance system. Eng Appl Artif Intell 67:226–234
83. Kong Y, Jia Y, Fu Y (2012) Learning human interaction by interactive phrases. In: Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part I 12, Springer, pp 300–313. https://doi.org/10.1007/978-3-642-33718-5_22
84 Köping L, Shirahama K, Grzegorzek M (2018) A general framework for sensor-based human activity recognition. Comp Biol Med 95:248–260

85. Kuehne H, Jhuang H, Garrote E, Poggio T, Serre T (2011) Hmdb: A large video database for human motion recognition. In: 2011 International Conference on Computer Vision, pp 2556–2563. https://doi.org/10.1109/ICCV.2011.6126543
86. Laptev I, Marsza-lek M, Schmid C, Rozenfeld B (2008) Learning realistic human actions from movies. In: IEEE Conference on Computer Vision & Pattern Recognition. https://doi.org/10.1109/CVPR.2008.4587756
87. Lazaridis L, Dimou A, Daras P (2018) Abnormal behavior detection in crowded scenes using density heatmaps and optical flow. In: 2018 26th European Signal Processing Conference (EUSIPCO), IEEE, pp 2060–2064. https://doi.org/10.23919/EUSIPCO.2018.8553620
88. Li C, Hou Y, Wang P, Li W (2017) Joint distance maps based action recognition with convolutional neural networks. IEEE Signal Process Lett 24(5):624–628
89. Li J, Han L, Zhang C, Li Q, Liu Z (2023) Spherical convolution empowered viewport prediction in 360 video multicast with limited fov feedback. ACM Trans Multimedia Comput Commun Appl 19(1). https://doi.org/10.1145/3511603
90. Li J, Zhang C, Liu Z, Hong R, Hu H (2023) Optimal volumetric video stream- ing with hybrid saliency based tiling. IEEE Trans Multimedia pp 2939–2953. https://doi.org/10.1109/TMM.2022.3153208
91. Li X, Xu Y et al (2022) Role of human-computer interaction healthcare sys- tem in the teaching of physiology and medicine. Comput Intel Neurosc 2022. https://doi.org/10.1155/2022/5849736
92. Li Y, Yang G, Su Z, Li S, Wang Y (2023) Human activity recognition based on multienvironment sensor data. Information Fusion 91:47–63
93. Lim KS, Ang KM, Isa NAM, Tiang SS, Rahman H, Chandrasekar B, Hussin EE, Lim WH (2023) Optimized machine learning model with modified par- ticle swarm optimization for data classification. In: Advances in Intelligent Manufacturing and Mechatronics: Selected Articles from the Innovative Man- ufacturing, Mechatronics & Materials Forum (iM3F 2022), Pahang, Malaysia, Springer Nature Singapore Singapore, pp 211–223. https://link.springer.com/chapter/10.1007/978-981-19-8703-8_18
94. Lin J, Gan C, Han S (2019) Tsm: Temporal shift module for efficient video understanding. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 7083–7093. https://doi.org/10.1109/ICCV.2019.00718
95. Liu C, Ying J, Han F, Ruan M (2018) Abnormal human activity recognition using bayes classifier and convolutional neural network. In: 2018 IEEE 3rd international conference on signal and image processing (ICSIP), IEEE, pp 33–37. https://doi.org/10.1109/SIPROCESS.2018.8600483
96. Liu H, Yuan H, Liu Q, Hou J, Zeng H, Kwong S (2022) A hybrid compression framework for color attributes of static 3d point clouds. IEEE Trans Circuits Syst Video Technol 32(3):1564–1577. https://doi.org/10.1109/TCSVT.2021.3069838
97. Liu Q, Yuan H, Hamzaoui R, Su H, Hou J, Yang H (2021) Reduced reference perceptual quality model with application to rate control for video-based point cloud compression. IEEE Trans Image Process 30:6623–6636. https://doi.org/10.1109/TIP.2021.3096060
98. Liu Y, Huang W, Jiang S, Zhao B, Wang S, Wang S, Zhang Y (2023) Transtm: A device-free method based on time-streaming multiscale transformer for human activity recognition. Defence Technology. https://doi.org/10.1016/j.dt.2023.02.021
99. Lobanova V, Bezdetnyy D, Anishchenko L (2023) Human activity recognition based on radar and video surveillance sensor fusion. In: 2023 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT), pp 025–028. https://doi.org/10.1109/USBEREIT58508.2023.10158846
100. Ma H, Li W, Zhang X, Gao S, Lu S (2019) Attnsense: Multi-level attention mechanism for multimodal human activity recognition. In: IJCAI, pp 3109– 3115. https://doi.org/10.24963/ijcai.2019%2F431
101. Mahdi MS, Mohammed AJ et al (2021) Detection of unusual activity in surveillance video scenes based on deep learning strategies. J Al-Qadisiyah Comp Sci Math 13(4):1
102. Manzi A, Dario P, Cavallo F (2017) A human activity recognition system based on dynamic clustering of skeleton data. Sensors 17(5):1100
103. Miao Y, Wang X, Wang S, Li R (2023) Adaptive switching control based on dynamic zero-moment point for versatile hip exoskeleton under hybrid lo- comotion. IEEE Trans Industr Electron 70(11):11443–11452. https://doi.org/10.1109/TIE.2022.3229343
104. Mishra A, Pandey A, Murthy HA (2020) Zero-shot learning for action recognition using synthesized features. Neurocomputing 390:117–130
105. Mohan A, Choksi M, Zaveri MA (2019) Anomaly and activity recognition using machine learning approach for video based surveillance. In: 2019 10th International Conference on Computing,

Communication and Networking Technologies (ICCCNT), IEEE, pp 1–6. https://doi.org/10.1109/icccnt45670.2019.8944396
106. Morales Garc´ıa S, Henao Baena C, Calvo Salcedo A (2023) Human activities recognition using semi-supervised svm and hidden markov models. TecnoL 26(56). https://doi.org/10.22430/22565337.2474
107. Moshe B (2005) Actions as space-time shapes. In: Proc. Tenth IEEE International Conference on Computer Vision, vol 2, pp 1395–1402. https://doi.org/10.1109/ICCV.2005.28
108. Moukafih Y, Hafidi H, Ghogho M (2019) Aggressive driving detection using deep learning-based time series classification. In: 2019 IEEE international symposium on INnovations in intelligent SysTems and applications (INISTA), IEEE, pp 1–5. https://doi.org/10.1109/INISTA.2019.8778416
109. Moya Rueda F, Grzeszick R, Fink GA, Feldhorst S, Ten Hompel M (2018) Convolutional neural networks for human activity recognition using body- worn sensors. In: Informatics, MDPI, vol 5, p 26. https://doi.org/10.3390/INFORMATICS5020026
110. Mukherjee S, Anvitha L, Lahari TM (2020) Human activity recognition in rgb-d videos by dynamic images. Multimed Tools Appl 79(27–28):19787–19801
111. Nadeem A, Jalal A, Kim K (2020) Human actions tracking and recognition based on body parts detection via artificial neural network. In: 2020 3rd International Conference on Advancements in Computational Sciences (ICACS), pp 1–6. https://doi.org/10.1109/ICACS47775.2020.9055951
112. Nagendran A, Harper D, Shah M (2010) New system performs persistent wide-area aerial surveillance. SPIE Newsroom 5:20–28
113. Najeh H, Lohr C, Leduc B (2022) Dynamic segmentation of sensor events for real-time human activity recognition in a smart home context. Sensors 22(14):5458
114. Najeh H, Lohr C, Leduc B (2022) Towards supervised real-time human activity recognition on embedded equipment. In: 2022 IEEE International Work- shop on Metrology for Living Environment (MetroLivEn), IEEE, pp 54–59. https://doi.org/10.1109/metrolivenv54405.2022.9826937
115. Nayak R, Pati UC, Das SK (2021) A comprehensive review on deep learning- based methods for video anomaly detection. Image Vis Comput 106:104078
116. Niebles JC, Chen CW, Fei-Fei L (2010) Modeling temporal structure of decomposable motion segments for activity classification. In: Computer Vision– ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5–11, 2010, Proceedings, Part II 11, Springer, pp 392–405. https://doi.org/10.1007/978-3-642-15552-9_29
117. Noori FM, Wallace B, Uddin MZ, Torresen J (2019) A robust human activity recognition approach using openpose, motion features, and deep recurrent neural network. In: Scandinavian conference on image analysis, Springer, pp 299–310. https://doi.org/10.1007/978-3-030-20205-7_25
118. Novopoltsev M, Verkhovtsev L, Murtazin R, Milevich D, Zemtsova I (2023) Fine-tuning of sign language recognition models: A technical report. arXiv preprint arXiv:230207693. https://doi.org/10.48550/arXiv.2302.07693
119. Nweke HF, Teh YW, Al-Garadi MA, Alo UR (2018) Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. Expert Syst Appl 105:233–261
120. Oh S, Hoogs A, Perera A, Cuntoor N, Chen CC, Lee JT, Mukherjee S, Aggarwal JK, Lee H, Davis L, Swears E, Wang X, Ji Q, Reddy K, Shah M, Vondrick C, Pirsiavash H, Ramanan D, Yuen J, Torralba A, Song B, Fong A, Roy-Chowdhury A, Desai M (2011) A large-scale benchmark dataset for event recognition in surveillance video. CVPR 2011:3153–3160. https://doi.org/10.1109/CVPR.2011.5995586
121. Pan S, Xu GJW, Guo K, Park SH, Ding H (2023) Video-based engagement estimation of game streamers: An interpretable multimodal neural network approach. IEEE Trans Games. https://doi.org/10.1109/TG.2023.3348230
122. Patil CM, Jagadeesh B, Meghana MN (2017) An approach of understanding human activity recognition and detection for video surveillance using hog descriptor and svm classifier. In: 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), pp 481–485. https://doi.org/10.1109/CTCEEC.2017.8455046
123. Paula LPO, Faruqui N, Mahmud I, Whaiduzzaman M, Hawkinson EC, Trivedi S (2023) A novel front door security (fds) algorithm using googlenet- bilstm hybridization. IEEE Access 11:19122–19134. https://doi.org/10.1109/ACCESS.2023.3248509
124. Peng JJ, Chen XG, Wang XK, Wang JQ, Long QQ, Yin LJ (2023) Pic- ture fuzzy decision-making theories and methodologies: a systematic re- view. Int J Syst Sci 54(13):2663–2675. https://doi.org/10.1080/00207721.2023.2241961
125. Perera AG, Law YW, Chahl J (2019) Drone-action: An outdoor recorded drone video dataset for action recognition. Drones 3(4):82

126. Pervaiz M, Jalal A (2023) Artificial neural network for human object interaction system over aerial images. In: 2023 4th International Conference on Advancements in Computational Sciences (ICACS), pp 1–6. https://doi.org/10.1109/ICACS55311.2023.10089722
127. Pervaiz M, Jalal A, Kim K (2021) Hybrid algorithm for multi people counting and tracking for smart surveillance. In: 2021 International Bhurban Conference on applied sciences and technologies (IBCAST), IEEE, pp 530–535
128. Piergiovanni A, Ryoo M (2020) Avid dataset: Anonymized videos from diverse countries. Adv Neural Inf Process Syst 33:16711–16721
129. Qi F, Tan X, Zhang Z, Chen M, Xie Y, Ma L (2024) Glass makes blurs: Learning the visual blurriness for glass surface detection. IEEE Trans Industr Inf 20(4):6631–6641. https://doi.org/10.1109/TII.2024.3352232
130. Qin Z, Liu H, Song B, Alazab M, Kumar PM (2021) Detecting and preventing criminal activities in shopping malls using massive video surveillance based on deep learning models. Annals of Operations Research pp 1–18. https://doi.org/10.1007/s10479-021-04264-0
131. Rahman H, Bukht TFN, Imran A, Tariq J, Tu S, Alzahrani A (2022) A deep learning approach for liver and tumor segmentation in ct images using resunet. Bioengineering 9(8):368
132. Rahman H, Naik Bukht TF, Ahmad R, Almadhor A, Javed AR et al (2023) Efficient breast cancer diagnosis from complex mammographic images using deep convolutional neural network. Comput Intel Neurosc 2023. https://doi.org/10.1155/2023%2F7717712
133. Rajpurkar OM, Kamble SS, Nandagiri JP, Nimkar AV (2020) Alert generation on detection of suspicious activity using transfer learning. In: 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, pp 1–7. https://doi.org/10.1109/ICCCNT49239.2020.9225263
134. Ray A, Kolekar MH, Balasubramanian R, Hafiane A (2023) Transfer learning enhanced vision-based human activity recognition: a decade-long analysis. Int J Inf Manag Data Insights 3(1):100142
135. Reddy KK, Shah M (2013) Recognizing 50 human action categories of web videos. Mach Vis Appl 24(5):971–981
136. Reinolds F, Neto C, Machado J (2022) Deep learning for activity recognition using audio and video. Electronics 11(5):782
137. Rodriguez MD, Ahmed J, Shah M (2008) Action mach a spatio-temporal maximum average correlation height filter for action recognition. In: 2008 IEEE conference on computer vision and pattern recognition, IEEE, pp 1–8. https://doi.org/10.1109/CVPR.2008.4587727
138. Ryoo MS, Chen CC, Aggarwal J, Roy-Chowdhury A (2010) An overview of contest on semantic description of human activities (sdha) 2010. Recogniz- ing Patterns in Signals, Speech, Images and Videos: ICPR 2010 Contests, Istanbul, Turkey, August 23–26, 2010, Contest Reports pp 270–285. https://doi.org/10.1007/978-3-642-17711-8_28
139. Ryumin D, Ivanko D, Ryumina E (2023) Audio-visual speech and gesture recognition by sensors of mobile devices. Sensors 23(4). https://doi.org/10.3390/s23042284, URL https://www.mdpi.com/1424-8220/23/4/2284
140. Saha SS, Sandha SS, Srivastava M (2021) Deep convolutional bidirectional lstm for complex activity recognition with missing data. Human Activity Recognition Challenge pp 39–53. https://doi.org/10.1007/978-981-15-8269-1_4
141. Saleem G, Bajwa UI, Raza RH (2023) Toward human activity recognition: a survey. Neural Comput Appl 35(5):4145–4182
142. Salleh S, Mahmud R, Rahman H, Yasiran SS (2017) Speed up robust features (surf) with principal component analysis-support vector machine (pca-svm) for benign and malignant classifications. J Fundam Appl Sci 9(5S):624–643
143. Samir H, Abd El Munim HE, Aly G (2018) Suspicious human activity recognition using statistical features. In: 2018 13th International Conference on Computer Engineering and Systems (ICCES), pp 589–594. https://doi.org/10.1109/ICCES.2018.8639457
144. Sanal Kumar K, Bhavani R (2019) Human activity recognition in egocentric video using pnn, svm, knn and svm+ knn classifiers. Clust Comput 22(Suppl 5):10577–10586
145. Sawanglok T, Thampairoj T, Songmuang P (2018) Activity recognition using kinect and comparison of supervised learning models for activity classification. In: 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), pp 1–6. https://doi.org/10.1109/iSAI-NLP.2018.8692801
146. Schuldt C, Laptev I, Caputo B (2004) Recognizing human actions: a local svm approach. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., IEEE, vol 3, pp 32–36. https://doi.org/10.1109/ICPR.2004.747

147. Shao J, Kang K, Change Loy C, Wang X (2015) Deeply learned attributes for crowded scene understanding. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4657–4666. https://doi.org/10.1109/CVPR.2015.7299097
148. Sharif M, Khan MA, Zahid F, Shah JH, Akram T (2020) Human action recognition: a framework of statistical weighted segmentation and rank correlation- based selection. Pattern Anal Appl 23:281–294
149. Shehzed A, Jalal A, Kim K (2019) Multi-person tracking in smart surveillance system for crowd counting and normal/abnormal events detection. In: 2019 International Conference on Applied and Engineering Mathemat- ics (ICAEM), pp 163–168. https://doi.org/10.1109/ICAEM.2019.8853756
150. Shelke S, Aksanli B (2019) Static and dynamic activity detection with am- bient sensors in smart spaces. Sensors 19(4):804
151. Shu X, Zhang L, Sun Y, Tang J (2020) Host–parasite: Graph lstm-in-lstm for group activity recognition. IEEE Trans Neural Netw Learn Syst 32(2):663–674
152. Sigurdsson GA, Varol G, Wang X, Farhadi A, Laptev I, Gupta A (2016) Hollywood in homes: Crowdsourcing data collection for activity understanding. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, Springer, pp 510–526. https://doi.org/10.1007/978-3-319-46448-0_31
153. Sigurdsson GA, Gupta A, Schmid C, Farhadi A, Alahari K (2018) Actor and observer: Joint modeling of first and third-person videos. In: proceedings of the IEEE conference on computer vision and pattern recognition, pp 7396– 7404. https://doi.org/10.1145/3265987.3265995
154. Singh S, Velastin SA, Ragheb H (2010) Muhavi: A multicamera human action video dataset for the evaluation of action recognition methods. In: 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, IEEE, pp 48–55. https://doi.org/10.1109/AVSS.2010.63
155. Singh T, Vishwakarma DK (2021) A deeply coupled convnet for human ac- tivity recognition using dynamic and rgb images. Neural Comput Appl 33:469–485
156. Smaira L, Carreira J, Noland E, Clancy E, Wu A, Zisserman A (2020) A short note on the kinetics-700–2020 human action dataset. arXiv preprint arXiv:201010864. https://doi.org/10.48550/arXiv.1907.06987
157. Soomro K, Zamir AR, Shah M (2012) Ucf101: A dataset of 101 human actions classes from videos in the wild. arXiv preprint arXiv:12120402. https://doi.org/10.48550/arXiv.1212.0402
158. Soomro K, Idrees H, Shah M (2016) Predicting the where and what of actors and actions through online action localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2648–2657. https://doi.org/10.1109/CVPR.2016.290
159. Sorkun MC, Dani¸sman AE, I˙ncel (2018) Human activity recognition with mobile phone sensors: Impact of sensors and window size. In: 2018 26th Signal Processing and Communications Applications Conference (SIU), pp 1–4. https://doi.org/10.1109/SIU.2018.8404569
160. Subasi A, Radhwan M, Kurdi R, Khateeb K (2018) Iot based mobile health- care system for human activity recognition. In: 2018 15th Learning and Technology Conference (LT), pp 29–34. https://doi.org/10.1109/LT.2018.8368507
161. Sukor AA, Rahim NA (2018) Activity recognition using accelerometer sensor and machine learning classifiers. In: 2018 IEEE 14th international colloquium on signal processing & its applications (CSPA), IEEE, pp 233–238. https://doi.org/10.1109/CSPA.2018.8368718
162. Sultani W, Shah M (2021) Human action recognition in drone videos using a few aerial training examples. Comput Vis Image Und p 103186. https://doi.org/10.1016/j.cviu.2021.103186
163. Sultani W, Chen C, Shah M (2018) Real-world anomaly detection in surveillance videos. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 6479–6488. https://doi.org/10.1109/CVPR.2018.00678
164. Sun Y, Peng Z, Hu J, Ghosh BK (2024) Event-triggered critic learning impedance control of lower limb exoskeleton robots in interactive environ- ments. Neurocomputing 564:126963. https://doi.org/10.1016/j.ne
165. Tang Y, Zhang L, Wu H, He J, Song A (2022) Dual-branch interac- tive networks on multichannel time series for human activity recognition. IEEE J Biomed Health Inform 26(10):5223–5234. https://doi.org/10.1109/JBHI.2022.3193148
166. Taylor W, Shah SA, Dashtipour K, Zahid A, Abbasi QH, Imran MA (2020) An intelligent non-invasive real-time human activity recognition system for next-generation healthcare. Sensors 20(9):2653
167. Uddin MZ, Hassan MM (2018) Activity recognition for cognitive assistance using body sensors data and deep convolutional neural network. IEEE Sens J 19(19):8413–8419

168. Ullah A, Muhammad K, Ding W, Palade V, Haq IU, Baik SW (2021) Efficient activity recognition using lightweight cnn and ds-gru network for surveillance applications. Appl Soft Comput 103:107102
169. Ullah A, Muhammad K, Hussain T, Baik SW (2021) Conflux lstms network: A novel approach for multi-view action recognition. Neurocomputing 435:321–329
170. Vaishnavi M, Sowmya J, Yaswanth M, Maruvarasi P (2023) Implementation of abnormal event detection using automated surveillance system. In: 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), IEEE, pp 1–6. https://doi.org/10.1109/ICCMC56507.2023.10084214
171. Vernikos I, Mathe E, Spyrou E, Mitsou A, Giannakopoulos T, Mylonas P (2019) Fusing handcrafted and contextual features for human activity recognition. In: 2019 14th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), IEEE, pp 1–6. https://doi.org/10.1109/SMAP.2019.8864848
172. Wang F, Ma M, Zhang X (2024) Study on a portable electrode used to detect the fatigue of tower crane drivers in real construction environment. IEEE Trans Instrum Meas 73. https://doi.org/10.1109/TIM.2024.3353274
173. Wang J, Chen Y, Hao S, Peng X, Hu L (2019) Deep learning for sensor-based activity recognition: A survey. Pattern Recogn Lett 119:3–11
174. Wang K, Williams H, Qian Z, Wei G, Xiu H, Chen W, Ren L (2023) Design and evaluation of a smooth-locking-based customizable prosthetic knee joint. J Mech Robot 16(4). https://doi.org/10.1115/1.4062498
175. Wang L, Qiao Y, Tang X (2015) Action recognition with trajectory-pooled deep-convolutional descriptors. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4305–4314. https://doi.org/10.1109/CVPR.2015.7299059
176. Wang Y, Qi Z, Li X, Liu J, Meng X, Meng L (2023) Multi-channel attentive weighting of visual frames for multimodal video classification. In: 2023 International Joint Conference on Neural Networks (IJCNN), IEEE, pp 1–8. https://doi.org/10.1109/IJCNN54540.2023.10192036
177. Weinland D, Ronfard R, Boyer E (2006) Free viewpoint action recognition using motion history volumes. Comput Vis Image Underst 104(2–3):249–257
178. Weinzaepfel P, Martin X, Schmid C (2016) Human action localization with sparse spatial supervision. arXiv preprint arXiv:160505197. https://doi.org/10.48550/arXiv.1605.05197
179. Welsh BC, Piza EL, Thomas AL, Farrington DP (2020) Private security and closed-circuit television (cctv) surveillance: A systematic review of function and performance. J Contemp Crim Justice 36(1):56–69
180. Wu Z, Zhu H, He L, Zhao Q, Shi J, Wu W (2023) Real-time stereo matching with high accuracy via spatial attention-guided upsampling. Appl Intell 53(20):24253–24274. https://doi.org/10.1007/s10489-023-04646-w
181. Xu J, Xiao L, López AM (2019) Self-supervised domain adaptation for computer vision tasks. IEEE Access 7:156694–156706
182. Xu J, Zhou G, Su S, Cao Q, Tian Z (2022) The development of a rigorous model for bathymetric mapping from multispectral satellite-images. Remote Sens 14(10). https://doi.org/10.3390/rs14102495
183. Xu M, Gao M, Chen YT, Davis LS, Crandall DJ (2019) Temporal recurrent networks for online action detection. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 5532–5541. https://doi.org/10.1109/ICCV.2019.00563
184. Xu W, Miao Z, Zhang XP, Tian Y (2017) A hierarchical spatio-temporal model for human activity recognition. IEEE Trans Multimedia 19(7):1494–1509
185. Yadav SK, Tiwari K, Pandey HM, Akbar SA (2021) A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions. Knowl-Based Syst 223:106970
186. Yan P, Li G, Xie Y, Li Z, Wang C, Chen T, Lin L (2019) Semi-supervised video salient object detection using pseudo-labels. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 7284–7293. https://doi.org/10.1109/ICCV.2019.00738
187. Yang D, Cui Z, Sheng H, Chen R, Cong R, Wang S, Xiong Z (2023) An occlusion and noise-aware stereo framework based on light field imaging for robust disparity estimation. IEEE Trans Comput. https://doi.org/10.1109/TC.2023.3343098
188. Yang H, Wen X, Geng Y, Wang Y, Wang X, Lu C (2022) Mpja-had: A multi- position joint angle dataset for human activity recognition using wearable sensors. In: 2022 International Conference on Advanced Mechatronic Systems (ICAMechS), pp 178–182. https://doi.org/10.1109/ICAMechS57222.2022.10003441

189. Yang Y, Zhan DC, Sheng XR, Jiang Y (2018) Semi-supervised multimodal learning with incomplete modalities. In: Proceedings of the Twenty- Seventh International Joint Conference on Artificial Intelligence, IJCAI-18, International Joint Conferences on Artificial Intelligence Organization, pp 2998–3004. https://doi.org/10.24963/ijcai.2018/416
190. Yang Y, Zhou DW, Zhan DC, Xiong H, Jiang Y (2019) Adaptive deep models for incremental learning: Considering capacity scalability and sustainability. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Association for Computing Machinery, New York, NY, USA, KDD '19, p 74–82. https://doi.org/10.1145/3292500.3330865
191. Yang Y, Fu ZY, Zhan DC, Liu ZB, Jiang Y (2021) Semi-supervised multimodal multi-instance multi-label deep network with optimal transport. IEEE Trans Knowl Data Eng 33(2):696–709. https://doi.org/10.1109/TKDE.2019.2932666
192. Yang Y, Zhang C, Xu Y, Yu D, chuan Zhan D, Yang J (2021) Rethink- ing label-wise cross-modal retrieval from a semantic sharing perspective. In: International Joint Conference on Artificial Intelligence. https://api.semanticscholar.org/CorpusID:237100828 or https://doi.org/10.24963/ijcai.2021%2F454
193. Yeung S, Russakovsky O, Jin N, Andriluka M, Mori G, Fei-Fei L (2018) Every moment counts: Dense detailed labeling of actions in complex videos. Int J Comput Vision 126:375–389
194. Yi C, Feng X et al (2021) Home interactive elderly care two-way video healthcare system design. J Healthcare Eng 2021. https://doi.org/10.1155/2021%2F6693617
195. Yimin D, Fudong C, Jinping L, Wei C (2019) Abnormal behavior detection based on optical flow trajectory of human joint points. In: 2019 Chinese Control And Decision Conference (CCDC), IEEE, pp 653–658. https://doi.org/10.1109/CCDC.2019.8833188
196. Yu L, Qian Y, Liu W, Hauptmann AG (2022) Argus++: Robust real-time activity detection for unconstrained video streams with overlapping cube pro- posals. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp 112–121. https://doi.org/10.1109/WACVW54805.2022.00017
197. Zeng M, Nguyen LT, Yu B, Mengshoel OJ, Zhu J, Wu P, Zhang J (2014) Convolutional neural networks for human activity recognition using mobile sensors. In: 6th international conference on mobile computing, applications and services, IEEE, pp 197–205. https://doi.org/10.4108/ICST.MOBICASE.2014.257786
198. Zhang J, Wu C, Wang Y, Wang P (2019) Detection of abnormal behavior in narrow scene with perspective distortion. Mach Vis Appl 30:987–998
199. Zhang R, Li L, Zhang Q, Zhang J, Xu L, Zhang B, Wang B (2023) Differential feature awareness network within antagonistic learning for infrared-visible ob- ject detection. IEEE Trans Circ Syst Video Tech. https://doi.org/10.1109/TCSVT.2023.3289142
200. Zhang Y, Po LM, Liu M, Rehman YAU, Ou W, Zhao Y (2020) Data-level information enhancement: Motion-patch-based siamese convolutional neural networks for human activity recognition in videos. Expert Syst Appl 147:113203
201. Zhao S, Liang W, Wang K, Ren L, Qian Z, Chen G, Ren L (2024) A multiaxial bionic ankle based on series elastic actuation with a parallel spring. IEEE Trans Industr Electron 71(7):7498–7510. https://doi.org/10.1109/TIE.2023.3310041
202. Zhao Y, Chen S, Liu S, Hu Z, Xia J (2024) Hierarchical equalization loss for long-tailed instance segmentation. IEEE Trans Multimedia 26:6943–6955. https://doi.org/10.1109/TMM.2024.3358080
203. Zhou G, Liu X (2022) Orthorectification model for extra-length linear array imagery. IEEE Trans Geosci Remote Sens 60. https://doi.org/10.1109/TGRS.2022.3223911
204. Zhou G, Tang Y, Zhang W, Liu W, Jiang Y, Gao E, Bai Y (2023) Shadow detection on high-resolution digital orthophoto map using semantic matching. IEEE Trans Geosci Remote Sens 61. https://doi.org/10.1109/TGRS.2023.3294531
205. Zhou L, Sun X, Zhang C, Cao L, Li Y (2024) Lidar-based 3-d glass detection and reconstruction in indoor environment. IEEE Trans Instrum Meas 73:1–11. https://doi.org/10.1109/TIM.2024.3375965
206. Zhou P, Qi J, Duan A, Huo S, Wu Z, Navarro-Alarcon D (2024) Imitating tool-based garment folding from a single visual observation using hand-object graph dynamics. IEEE Trans Industr Inf 20(4):6245–6256. https://doi.org/10.1109/TII.2023.3342895
207. Zhou Y et al (2022) Construction of a digital elderly care service system based on human-computer interaction from the perspective of smart elderly care. Comput Intel Neurosc 2022. https://doi.org/10.1155/2022%2F1500339
208. Zhou Z, Wang Y, Zhou G, Nam K, Ji Z, Yin C (2023) A twisted gaussian risk model considering target vehicle longitudinal-lateral motion states for host vehicle trajectory planning. IEEE Trans Intell Transp Syst 24(12):13685–21397. https://doi.org/10.1109/TITS.2023.3298110

209. Zolfaghari M, Singh K, Brox T (2018) Eco: efficient convolutional network for online video understanding. In: Proceedings of the European conference on computer vision (ECCV), pp 695–712. https://doi.org/10.1007/978-3-030-01216-8_43
210. Zolfaghari S, Keyvanpour MR, Zall R (2017) Analytical review on ontological human activity recognition approaches. Int J Bus Res (IJEBR) 13(2):58–78

Authors and Affiliations

Tanvir Fatima Naik Bukht[1] · Hameedur Rahman[1] · Momina Shaheen[2] · Asaad Algarni[3] · Nouf Abdullah Almujally[4] · Ahmad Jalal[1]

✉ Hameedur Rahman
hameed.rahman@mail.au.edu.pk

✉ Momina Shaheen
momina.shaheen@roehampton.ac.uk

Tanvir Fatima Naik Bukht
211893@students.au.edu.pk; tanvir.fatima@au.edu.pk

Asaad Algarni
Asaad.Algarni@nbu.edu.sa

Nouf Abdullah Almujally
naalmujally@pnu.edu.sa

Ahmad Jalal
ahmadjalal@mail.au.edu.pk

[1] Faculty of Computing and AI, Air University, E-9, Islamabad, Pakistan

[2] Department of Computing, School of Arts, Humanities and Social Sciences, University of Roehampton London, UK SW15 5PJ, London, UK

[3] Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, 91911 Rafha, Saudi Arabia

[4] Department of Information Systems, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, 11671 Riyadh, Saudi Arabia